

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: November 17, 2004, 14:58:39 ; Search time 98 Seconds  
(without alignments)  
530.773 Million cell updates/sec

Title: US-10-705-716A-4  
Perfect score: 778  
Sequence: 1 MGCGSRADAEPRYIESWT.....VTDSIQQMDRSRRITKNCVN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	778	100.0	145	4	AAB95018 Human pro
2	778	100.0	145	5	Aao19498 Hs1 prote
3	778	100.0	145	6	Abr58646 Human can
4	778	100.0	145	7	Adc31800 Human nov
5	778	100.0	145	7	Adm46959 Brain and
6	778	100.0	145	8	Ado48475 Human PTH
7	750.5	96.5	180	7	Adm46961 Brain and
8	649	83.4	145	8	Ado48479 Mouse PTH
9	645	82.9	145	8	Ado48473 Rat PTH r
10	569.5	73.2	149	7	Adm46963 Brain and
11	405.5	52.1	92	5	Adq81902 Human dio
12	309	39.7	73	7	Adm46962 Brain and
13	299	38.4	54	7	Adm46960 Brain and
14	299	38.4	54	8	Ado48477 Human PTH
15	293	37.7	80	7	Adm46964 Brain and
16	278	35.7	54	8	Ado48481 Mouse PTH
17	140	18.0	25	7	Adm46979 Brain and
18	99	12.7	18	7	Adm46977 Brain and
19	92.5	11.9	1001	5	AAU98903 Rat inosi
20	91.5	11.8	210	3	Agi5390 Arabidops
21	91.5	11.8	222	3	Agi5389 Arabidops
22	89	11.4	16	8	Ado48482 PTH respo
23	89	11.4	321	4	Abb71629 Drosophil
24	88	11.3	298	5	Aau87086 Siglec-BM
25	87.5	11.2	369	3	Aay71485 Human MAG

## ALIGNMENTS

RESULT 1  
AAB95018  
ID AAB95018 standard; protein; 145 AA.  
XX  
AC AAB95018;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:16726.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000BP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
XX  
PR 27-AUG-1999; 99JP-00300253.  
XX  
PR 11-JAN-2000; 2000JP-00118776.  
XX  
PR 02-MAY-2000; 2000JP-00183767.  
XX  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
(HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
PS Claim 8; SEQ ID NO 16726; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

26 87.5 11.2 369 4 AAB80297 Human pro  
27 87.5 11.2 369 6 ABR48215 Human bla  
28 87.5 11.2 369 6 ABU56516 Lung canc  
29 87.5 11.2 383 8 ABO58424 Human gen  
30 86.5 11.1 1192 4 ABG02038 Novel hum  
31 86 11.1 342 8 ADM87821 Human EST  
32 86 11.1 508 4 ABG09910 Novel hum  
33 86 11.1 586 5 ABB04711 Human PPI  
34 86 11.1 991 4 AAB83195 Human Rec  
35 86 11.1 991 8 ADO55153 Protein #  
36 85 10.9 486 6 ABU34190 Protein e  
37 84.5 10.9 1111 5 AAO17108 Murine Gl  
38 84 10.8 514 4 AAG81103 Mycobacte  
39 83.5 10.7 684 4 ABB69330 Drosophil  
40 83.5 10.7 899 4 ABB65489 Drosophil  
41 83.5 10.7 899 4 ABB65488 Drosophil  
42 83 10.7 659 4 ABB65280 Drosophil  
43 82.5 10.6 547 4 ABG14111 Novel hum  
44 82.5 10.6 1017 4 AAB59813 TutD prot  
45 82.5 10.6 1615 4 AAB59826 Protein #

CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 145 AA;

Query Match 100.0%; Score 778; DB 4; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-70;  
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIEPRYIESWTRETESTWLTYYTDSADAPPAAAPDSGPEAGGLHSGMLEDGL 60  
 DB 1 MCGGSRADAIEPRYIESWTRETESTWLTYYTDSADAPPAAAPDSGPEAGGLHSGMLEDGL 60  
 QY 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGGLTKQKGLQTTAKRDAKRMPEAK 120  
 DB 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGGLTKQKGLQTTAKRDAKRMPEAK 120  
 QY 121 EVTINVTDISIQQMDRSRRITKNCVN 145  
 DB 121 EVTINVTDISIQQMDRSRRITKNCVN 145

RESULT 2  
 AA019498  
 ID AA019498 standard; protein; 145 AA.  
 AC AA019498;  
 DT 20-DEC-2002 (first entry)  
 XX  
 DE HSI protein variant.  
 XX  
 KW HSI; variant; cancer; tumour; unigene cluster; cytostatic; metastasis;  
 KW EST; expressed sequence tag; colon cancer; stomach cancer; breast cancer;  
 KW HSI69395; HSI27144; HS2; HSI132793; HS3.  
 XX  
 OS Unidentified.

XX  
 XX DE10103694-A1.  
 XX  
 XX 01-AUG-2002.  
 XX  
 XX 26-JAN-2001; 2001DE-01003694.  
 XX  
 XX 26-JAN-2001; 2001DE-01003694.  
 XX  
 XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 XX  
 XX Brett D, Kemner W;  
 XX  
 XX WPI; 2002-644836/70.  
 XX  
 XX N-PSDB; AAL50100.

XX  
 PT Diagnosis and therapy of tumors, by determining expression rates of  
 PT specific expressed sequence tags of the unigene cluster, and subsequently  
 PT blocking their expression.  
 XX  
 XX Claim 10; Page 5; 10pp; German.  
 XX  
 XX The present invention relates to the use of expressed sequence tags  
 CC (ESTs), or variants, of the unigene cluster HSI69395 (HS1), HSI27144

CC (HS2) and/or HSI132793 (HS3) for diagnosis and therapy of tumours, in  
 CC which their expression rates in tumour cells and/or lymph nodes are  
 CC determined. The EST sequences are useful as prognostic markers of  
 CC survival of cancer patients (high levels of EST-related mRNA are  
 CC associated with a poor prognosis, specifically correlated with  
 CC development of metastases); and for diagnosis and/or therapy of solid  
 CC tumours, particularly of colon, stomach and breast. The present sequence  
 CC is a variant of the HSI1 protein shown in the exemplification of the  
 CC invention  
 XX  
 SQ Sequence 145 AA;

Query Match 100.0%; Score 778; DB 5; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-70;  
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIEPRYIESWTRETESTWLTYYTDSADAPPAAAPDSGPEAGGLHSGMLEDGL 60  
 DB 1 MCGGSRADAIEPRYIESWTRETESTWLTYYTDSADAPPAAAPDSGPEAGGLHSGMLEDGL 60  
 QY 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGGLTKQKGLQTTAKRDAKRMPEAK 120  
 DB 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGGLTKQKGLQTTAKRDAKRMPEAK 120  
 QY 121 EVTINVTDISIQQMDRSRRITKNCVN 145  
 DB 121 EVTINVTDISIQQMDRSRRITKNCVN 145

RESULT 3  
 ABR58646  
 ID ABR58646 standard; protein; 145 AA.  
 AC ABR58646;  
 XX  
 XX 09-JUL-2003 (first entry)  
 XX  
 DE Human cancer related protein SEQ ID NO:303.  
 XX  
 KW Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;  
 KW heart disease; atherosclerosis; endometriosis.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2003025138-A2.  
 XX  
 XX 27-MAR-2003.  
 XX  
 XX 17-SEP-2002; 2002WO-US029560.  
 XX  
 XX 17-SEP-2001; 2001US-0323469P.  
 XX  
 XX 20-SEP-2001; 2001US-0323887P.  
 XX  
 XX 13-NOV-2001; 2001US-0350666P.  
 XX  
 XX 08-FEB-2002; 2002US-0355145P.  
 XX  
 XX 08-FEB-2002; 2002US-0355257P.  
 XX  
 XX 12-APR-2002; 2002US-0372246P.  
 XX  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX  
 XX Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;  
 XX  
 XX Zlotnik A;  
 XX  
 XX WPI; 2003-354600/33.  
 XX  
 XX N-PSDB; ACC72796.

XX  
 PT New genes that are up-regulated or down-regulated in cancers, useful as  
 PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as  
 PT therapeutic targets for screening drugs for treating these diseases.  
 XX  
 XX Claim 12; Page 753; 767pp; English.  
 XX  
 XX The present invention describes an isolated nucleic acid molecule, which  
 CC comprises the sequence of any of the genes that are up-regulated or down-

regulated in specific cancers (e.g. about 1031 genes up-regulated in acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer related gene nucleotide sequences which encode the proteins given in ABR58521 to ABR58709. Also described: (1) determining the presence or absence of a pathological cell in a patient; (2) an expression vector comprising a nucleic acid molecule described above; (3) a host cell comprising the vector; (4) an isolated polypeptide, which is encoded by the nucleic acid; (5) an antibody that specifically binds the polypeptide of (4); (6) specifically targeting a compound to a pathological cell in a patient by administering to the patient the antibody above; and (7) a drug screening assay. The nucleic acid is useful as diagnostic markers or therapeutic targets. In particular, the nucleic acid is useful for diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow, bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary, pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases, atherosclerosis and endometriosis. The nucleic acid is also useful in drug screening, particularly for identifying agents for treating these pathologies

Query Match 100.0%; Score 778; DB 6; Length 145;  
Best Local Similarity 100.0%; Pred. No. 3.8e-70;  
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCGGSRADAIEPRYVESWTRETESTWLTYSDDAPPSSAAAPDSGPEAGLHSGMLEDGL 60  
DB 1 MCGGSRADAIEPRYVESWTRETESTWLTYSDDAPPSSAAAPDSGPEAGLHSGMLEDGL 60

QY 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGPLTKQKGLQTTAKRDAKRMMPAK 120  
DB 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGPLTKQKGLQTTAKRDAKRMMPAK 120

QY 121 EVTINVTDSIQOMDRSRRTKNCVN 145  
DB 121 EVTINVTDSIQOMDRSRRTKNCVN 145

RESULT 4  
ADC31800  
ID ADC31800 standard; protein; 145 AA.  
AC ADC31800;  
XX  
XX 18-DEC-2003 (first entry)  
XX Human novel polypeptide sequence, SEQ ID NO:1882.  
XX Human; diagnostic; drug screening; forensics; gene mapping;  
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KW ulcers; osteoporosis; autoimmune disease; cancer;  
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KW neuroprotective; anti-anaemic; anticoagulant; thrombolytic; vulnary;  
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
KW gene therapy; chromosome 8.

OS Homo sapiens.  
XX WO2003029271-A2.  
XX 10-APR-2003.  
XX 24-SEP-2002; 2002WO-US030474.  
XX 24-SEP-2001; 2001US-0324631P.  
XX (HYSE-) HYSEQ INC.  
XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;  
XX

DR WPI; 2003-371981/35.  
DR N-PSDB; ADC30829.  
XX New polynucleotide and polypeptide useful for diagnosing, preventing or treating conditions such as neurodegenerative diseases, anemias, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer.  
XX Claim 20; SEQ ID NO 1882; 1185pp; English.  
PS The invention relates to 971 novel human cDNA sequences (ADC29919-ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The invention also relates to nucleic acid sequences over 99% identical with the novel human cDNAs. The invention additionally encompasses expression vectors and host cells comprising a nucleic acid of the invention; the recombinant production of a polypeptide of the invention; an antibody against a polypeptide of the invention; a method of detecting polynucleotides or polypeptides of the invention; and methods of identifying a compound which binds to a polypeptide of the invention. The invention further discloses methods of preventing, treating or ameliorating a medical condition; kits comprising polynucleotide probes and/or monoclonal antibodies for carrying out the methods of the invention; methods for the identification of compounds that modulate the expression or activity of the polynucleotide and/or polypeptide; and 767 contig sequences corresponding to the cDNA sequences of the invention (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC33394). The nucleic acids and polypeptides of the invention are useful in diagnostics, drug screening, forensics, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a specifically claimed human polypeptide sequence of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 100.0%; Score 778; DB 7; Length 145;  
Best Local Similarity 100.0%; Pred. No. 3.8e-70;  
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCGGSRADAIEPRYVESWTRETESTWLTYSDDAPPSSAAAPDSGPEAGLHSGMLEDGL 60  
DB 1 MCGGSRADAIEPRYVESWTRETESTWLTYSDDAPPSSAAAPDSGPEAGLHSGMLEDGL 60

QY 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGPLTKQKGLQTTAKRDAKRMMPAK 120  
DB 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGPLTKQKGLQTTAKRDAKRMMPAK 120

QY 121 EVTINVTDSIQOMDRSRRTKNCVN 145  
DB 121 EVTINVTDSIQOMDRSRRTKNCVN 145

RESULT 5  
ADM46959  
ID ADM46959 standard; protein; 145 AA.  
XX ADM46959;  
XX 03-JUN-2004 (first entry)  
XX Brain and Acute Leukemia, Cytoplasmic alternate protein #1.  
XX acute myelogenous leukemia; gene expression; BAALC;  
KW

KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 XX Cytoplasmic; exon.  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FH Misc-difference 41  
 FT /note= "encoded by GCS"  
 FT

XX WO2003040347-A2.

XX 15-MAY-2003.

XX 12-NOV-2002; 2002WO-US036375.

XX 09-NOV-2001; 2001US-0348210P.

XX (OHIS ) UNIV OHIO STATE RES FOUND.

XX Tanner SM, De La Chapell A;

XX WPI; 2003-441564/41.

XX N-PSDB; ADM4651.

XX Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 XX in a patient comprises assaying for the overexpression of one or more  
 XX BAALC transcripts in cells obtained from the patient.

XX Disclosure; SEQ ID NO 17; 78pp; English.

XX The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
 CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.

XX Sequence 145 AA;

Query Match 100.0%; Score 778; DB 7; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-70;  
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCGGSRADALEPRYESTWETESTWLTYSDDAPPSAAAPDSGPAGGLHSGMLDGL 60

Db 1 MCGGSRADALEPRYESTWETESTWLTYSDDAPPSAAAPDSGPAGGLHSGMLDGL 60

Qy 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGGLTKQNGLQTTAKRDKRMPAK 120

Db 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGGLTKQNGLQTTAKRDKRMPAK 120

Qy 121 EVINVTDSIQMDRSRRITKNCVN 145

Db 121 EVINVTDSIQMDRSRRITKNCVN 145

RESULT 6  
 ADO48475  
 ID ADO48475 standard; protein; 145 AA.  
 XX  
 AC ADO48475;

XX 12-AUG-2004 (first entry)

XX Human PTH responsive gene protein.

XX PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
 XX transgenic animal; osteopathic; gene therapy; osteoporosis; human.  
 XX Homo sapiens.

XX

PN WO2004044152-A2.

XX 27-MAY-2004.

XX 10-NOV-2003; 2003WO-US035655.

XX 12-NOV-2002; 2002US-0425532P.

XX (AMHP ) WYETH.

XX Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;

XX WPI; 2004-420299/39.

XX N-PSDB; ADO48474.

XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in

XX preparing a composition for diagnosing, treating or preventing bone

XX related disorders, e.g., osteoporosis.

XX Claim 9; SEQ ID NO 4; 169pp; English.

XX The invention relates to a novel PTH responsive gene (PAIGB) fragment  
 CC encoding a polypeptide. The invention further comprises: a chimeric  
 CC construct comprising the isolated nucleic acid fragment operatively  
 CC linked to suitable regulatory sequences; a host cell transformed with the  
 CC chimeric construct; a vector comprising the nucleic acid fragment;  
 CC obtaining a nucleic acid fragment encoding the polypeptide; a method for  
 CC obtaining a polypeptide; detecting the presence of the nucleic acid  
 CC fragment; an antibody that specifically binds to one or more epitopes of  
 CC a PAIGB polypeptide; a composition for regulating bone-forming activity  
 CC in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
 CC ; an agent that alters the expression of PAIGB gene or polypeptide;  
 CC determining whether an agent alters the expression of PAIGB mRNA;  
 CC screening agents for effectiveness in altering expression of the nucleic  
 CC acid fragment; screening for agents useful for treating bone related  
 CC disorders; evaluating the efficacy of a treatment of a bone related  
 CC disorder in a subject; identifying polypeptides capable of binding to  
 CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
 CC related agent; a transgenic animal comprising the DNA; an animal model  
 CC for the study of bone density modulation comprising a first group of  
 CC animals composed of the transgenic animal and a second group of control  
 CC animals; studying bone mass determinants; studying the modulation of bone  
 CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
 CC for treating bone related disorders; identifying whether an agent which  
 CC has bone forming activity; and a stably transfected cell line comprising  
 CC two constructs, the first construct comprising a ligand binding domain  
 CC linked to a DNA binding domain which is linked to an activation domain  
 CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

XX Sequence 145 AA;

Query Match 100.0%; Score 778; DB 8; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-70;  
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCGGSRADALEPRYESTWETESTWLTYSDDAPPSAAAPDSGPAGGLHSGMLDGL 60

Db 1 MCGGSRADALEPRYESTWETESTWLTYSDDAPPSAAAPDSGPAGGLHSGMLDGL 60

Qy 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGGLTKQNGLQTTAKRDKRMPAK 120

Db 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSLSSGGLTKQNGLQTTAKRDKRMPAK 120

Qy 121 EVINVTDSIQMDRSRRITKNCVN 145



Db 121 EVTINVTDISIQQMDSRRITKNCVN 145  
|||||  
RESULT 7  
AD46961  
ID ADM46961 standard; protein; 180 AA.  
XX AC  
XX AD46961;  
XX  
XX 03-JUN-2004 (first entry)  
XX  
XX Brain and Acute Leukemia, Cytoplasmic alternate protein #3.  
XX DE  
XX acute myelogenous leukemia; gene expression; BAALC;  
XX KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
XX KW Cytoplasmic; exon.  
XX OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 41  
FT /note= "encoded by GCS"  
XX  
XX WO2003040347-A2.  
XX PN  
XX 15-MAY-2003.  
XX PD  
XX 12-NOV-2002; 2002WO-US036375.  
XX PF  
XX 09-NOV-2001; 2001US-0348210P.  
XX PR  
XX (OHIS ) UNIV OHIO STATE RES FOUND.  
XX PA  
XX Tanner SM, De La Chapell A;  
XX PI  
XX WPI; 2003-441564/41.  
XX DR N-PSDB; ADM46953.  
XX  
XX Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
PT in a patient comprises assaying for the overexpression of one or more  
PT BAALC transcripts in cells obtained from the patient.  
XX  
XX Disclosure; SEQ ID NO 19; 78pp; English.  
XX  
XX The invention relates to a method of characterizing acute myelogenous  
CC leukemia (AML) in a patient by assaying for the overexpression of one or  
CC more BAALC transcripts in cells obtained from the patient, where an  
CC overexpression indicates that the patient has an aggressive form of AML.  
CC The methods, kits and probes are useful for characterizing acute or  
CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
CC spliced RNA consisting of exons 1, 6 and 8.  
XX  
XX Sequence 180 AA;  
Query Match 96.5%; Score 750.5; DB 7; Length 180;  
Best Local Similarity 80.6%; Pred. No. 3e-67;  
Matches 145; Conservative 0; Mismatches 0; Indels 35; Gaps 1;  
QY 1 MCGGSRADAIPRYVESWTRTESTWLTWYTDSDAPPSAAAPDSGPAGGLHS----- 53  
Db 1 MCGGSRADAIPRYVESWTRTESTWLTWYTDSDAPPSAAAPDSGPAGGLHSVLEAKS 60  
QY 54 -----QWLEDGLPSNGVPRSTAPGGIPNPEKTKNCST 85  
Db 61 KIKAPTDSVSDGLFSASKWAPLAVFSHGMLEDGLPSNGVPRSTAPGGIPNPEKTKNCST 120  
QY 86 QCPNQSLSGGPLTQKQNGLOTTEAKRMPAKEVTINVTDISIQQMDSRRITKNCVN 145  
Db 121 QCPNQSLSGGPLTQKQNGLOTTEAKRMPAKEVTINVTDISIQQMDSRRITKNCVN 180

RESULT 8  
ADO48479  
ID ADO48479 standard; protein; 145 AA.  
XX AC  
XX ADO48479;  
XX  
XX 12-AUG-2004 (first entry)  
XX  
XX Mouse PTH responsive gene protein.  
XX  
XX PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
XX KW transgenic animal; osteopathic; gene therapy; osteoporosis; mouse;  
XX KW murine.  
XX OS Mus sp.  
XX PN  
XX WO2004044152-A2.  
XX PD  
XX 27-MAY-2004.  
XX PF 10-NOV-2003; 2003WO-US035655.  
XX PR 12-NOV-2002; 2002US-0425532P.  
XX PA (AMHP ) WYETH.  
XX PI Robinson JA, Stojanovic-Susulic V, Babić P, Murrills RJ;  
XX WPI; 2004-420299/39.  
XX DR N-PSDB; ADO48478.  
XX  
XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
PT preparing a composition for diagnosing, treating or preventing bone  
PT related disorders, e.g., osteoporosis.  
XX  
XX Claim 9; SEQ ID NO 8; 169pp; English.  
XX  
XX The invention relates to a novel PTH responsive gene (PAIGB) fragment  
CC encoding a polypeptide. The invention further comprises: a chimeric  
CC construct comprising the isolated nucleic acid fragment operatively  
CC linked to suitable regulatory sequences; a host cell transformed with the  
CC chimeric construct; a vector comprising the nucleic acid fragment;  
CC obtaining a nucleic acid fragment encoding the polypeptide; a method for  
CC obtaining a polypeptide; detecting the presence of the nucleic acid  
CC fragment; an antibody that specifically binds to one or more epitopes of  
CC a PAIGB polypeptide; a composition for regulating bone-forming activity  
CC in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
CC ; an agent that alters the expression of PAIGB gene or polypeptide;  
CC determining whether an agent alters the expression of PAIGB mRNA;  
CC screening agents for effectiveness in altering expression of the nucleic  
CC acid fragment; screening for agents useful for treating bone related  
CC disorders; evaluating the efficacy of a treatment of a bone related  
CC disorder in a subject; identifying polypeptides capable of binding to  
CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
CC related agent; a transgenic animal comprising the DNA, an animal model  
CC for the study of bone density modulation comprising a first group of  
CC animals composed of the transgenic animal and a second group of control  
CC animals; studying bone mass determinants; studying the modulation of bone  
CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
CC for treating bone related disorders; identifying whether an agent which  
CC has bone forming activity; and a stably transfected cell line comprising  
CC two constructs, the first construct comprising a ligand binding domain  
CC linked to a DNA binding domain which is linked to an activation domain  
CC all of which expression is driven by a constitutive promoter, the second  
CC construct comprising multiple copies of DNA binding elements linked to a  
CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
CC used to treat disorders by gene therapy. The nucleic acid is useful in  
CC preparing a composition for diagnosing, treating or preventing bone  
CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
CC responsive gene protein of the invention.

XX SQ Sequence 145 AA;

Query Match 83.4%; Score 649; DB 8; Length 145;  
 Best Local Similarity 83.4%; Pred. No. 3.8e-57;  
 Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MCGGSRADALEPRYYSWTRETESTWLTYSDDAPPSSAAPDGPAGGLHSGWLEDGL 60  
 Db 1 MCGGSRADALEPRYYSWTRETESTWLTYSDDAPPSSAAPDGPAGGLHSGWLEDGL 60

Qy 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNQSLSGGLTKQNGLOTTEAKRDARMPAK 120  
 Db 61 SSGVLRPAAPGGIANPEKKNKCCGTCPCNSQNLSSGLTKQNGLOTTEAKRDARMSAR 120

Qy 121 EVTINVTDSTQOMDRSRRTKNCVN 145  
 Db 121 EVAINVTENIRQMDRSKRVTNKCIN 145

RESULT 9  
 ADO48473  
 ID ADO48473 standard; protein; 145 AA.  
 AC ADO48473;  
 DT 12-AUG-2004 (first entry)  
 DE Rat PTH responsive gene protein.  
 XX PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
 KW transgenic animal; osteopathic; gene therapy; osteoporosis; rat.  
 XX Rattus sp.  
 XX WO2004044152-A2.  
 XX 27-MAY-2004.  
 XX 10-NOV-2003; 2003WO-US035655.  
 XX 12-NOV-2002; 2002US-0425532P.  
 XX (AMHP ) WYETH.  
 XX Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;  
 XX WPI; 2004-420299/39.  
 XX N-PSDB; ADO48472.  
 XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
 PT preparing a composition for diagnosing, treating or preventing bone  
 PT related disorders, e.g., osteoporosis.  
 XX Claim 9; SEQ ID NO 2; 169pp; English.

The invention relates to a novel PTH responsive gene (PAIGB) fragment  
 encoding a polypeptide. The invention further comprises: a chimeric  
 construct comprising the isolated nucleic acid fragment operatively  
 linked to suitable regulatory sequences; a host cell transformed with the  
 chimeric construct; a vector comprising the nucleic acid fragment;  
 obtaining a nucleic acid fragment encoding the polypeptide; a method for  
 obtaining a polypeptide; detecting the presence of the nucleic acid  
 fragment; an antibody that specifically binds to one or more epitopes of  
 a PAIGB polypeptide; a composition for regulating bone-forming activity  
 in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
 ; an agent that alters the expression of PAIGB gene or polypeptide;  
 determining whether an agent alters the expression of PAIGB gene or polypeptide;  
 screening agents for effectiveness in altering expression of the nucleic  
 acid fragment; screening for agents useful for treating bone related  
 disorders; evaluating the efficacy of a treatment of a bone related  
 disorder in a subject; identifying polypeptides capable of binding to  
 CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone

CC related agent; a transgenic animal comprising the DNA; an animal model  
 CC for the study of bone density modulation comprising a first group of  
 CC animals composed of the transgenic animal and a second group of control  
 CC animals; studying bone mass determinants; studying the modulation of bone  
 CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
 CC for treating bone related disorders; identifying whether an agent which  
 CC has bone forming activity; and a stably transfected cell line comprising  
 CC two constructs, the first construct comprising a ligand binding domain  
 CC linked to a DNA binding domain which is linked to an activation domain  
 CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

SQ Sequence 145 AA;

Query Match 82.9%; Score 645; DB 8; Length 145;  
 Best Local Similarity 83.4%; Pred. No. 9.6e-57;  
 Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MCGGSRADALEPRYYSWTRETESTWLTYSDDAPPSSAAPDGPAGGLHSGWLEDGL 60  
 Db 1 MCGGSRADALEPRYYSWTRETESTWLTYSDDAPPSSAAPDGPAGGLHSGWLEDGL 60

Qy 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNQSLSGGLTKQNGLOTTEAKRDARMPAK 120  
 Db 61 SSGVLRPAAPGGIANPEKKNKCCGTCPCNSQNLSSGLTKQNGLOTTEAKRDARMSAR 120

Qy 121 EVTINVTDSTQOMDRSRRTKNCVN 145  
 Db 121 EVAINVTENIRQMDRSKRVTNKCIN 145

RESULT 10

ADAM46963  
 ID ADM46963 standard; protein; 149 AA.  
 XX ADM46963;  
 AC ADM46963;  
 DT 03-JUN-2004 (first entry)  
 DE Brain and Acute Leukemia, Cytoplasmic alternate protein #5.  
 XX acute myelogenous leukemia; gene expression; BAALC;  
 KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 KW Cytoplasmic; exon.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 FT Misc-difference 41 /note= "encoded by GCS"  
 XX WO2003040347-A2.  
 XX 15-MAY-2003.  
 XX 12-NOV-2002; 2002WO-US036375.  
 XX 09-NOV-2001; 2001US-0348210P.  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Tanner SM, De La Chapell A;  
 XX WPI; 2003-441564/41.  
 XX N-PSDB; ADM46955.

PT Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
PT in a patient comprises assaying for the overexpression of one or more  
PT BAALC transcripts in cells obtained from the patient.

PS Disclosure; SEQ ID NO 21; 78pp; English.

XX The invention relates to a method of characterizing acute myelogenous  
XX leukemia (AML) in a patient by assaying for the overexpression of one or  
XX more BAALC transcripts in cells obtained from the patient, where an  
XX overexpression indicates that the patient has an aggressive form of AML.  
XX The methods, kits and probes are useful for characterizing acute or  
XX chronic myelogenous leukemia, or prostate cancer. They are also useful  
XX for detecting BAALC overexpression. This sequence corresponds to a BAALC  
XX (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
XX spliced RNA consisting of exons 1, 6 and 8.

XX Sequence 149 AA;

Query Match 73.2%; Score 569.5; DB 7; Length 149;  
Best Local Similarity 75.7%; Pred. No. 4e-49;  
Matches 109; Conservative 0; Mismatches 0; Indels 35; Gaps 1;

QY 1 MCGGSRADAIEPRYESTRETESTWLTYSDDAPPSSAAAPDSGPEAGGLHS----- 53

DB 1 MCGGSRADAIEPRYESTRETESTWLTYSDDAPPSSAAAPDSGPEAGGLHSVLEAEKS 60

QY 54 -----GMLEDGLPSNGVPRSTAPGGINPEKKTNCET 85

DB 61 KIKAPTDSVSDGLFSASKMAPLAVFSGHGMLEDGLPSNGVPRSTAPGGINPEKKTNCET 120

QY 86 QCPNPQSLSSGGLTKQKGLQTTTE 109

DB 121 QCPNPQSLSSGGLTKQKGLQTTTE 144

RESULT 11

ID ADQ81902 standard; protein; 92 AA.

AC ADQ81902;

DT 09-SEP-2004 (first entry)

DE Human dioxigenase 10.12.

XX Human; enzyme; dioxigenase 10.12; malignant tumour; inflammation;  
KW immunological disease; haemopathy; HIV infection.

OS Homo sapiens.

FN CN1344798-A.

PD 17-APR-2002.

XX 29-SEP-2000; 2000CN-00125495.

PR 29-SEP-2000; 2000CN-00125495.

XX (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.

PI Mao Y, Xie Y;

DR WPI; 2002-509506/55.

DR N-PSDB; ADQ81901.

XX New polypeptide human dioxigenase 10.12 and polynucleotides encoding this  
PT polypeptide, useful for treating various diseases, such as malignant  
PT tumors, inflammations, immunological diseases, hemopathy and HIV  
PT infection.

XX Claim 1; SEQ ID NO 2; 33pp; Chinese.

XX The present invention discloses a new kind of polypeptide, human

CC dioxigenase 10.12, polynucleotides encoding this polypeptide, a DNA  
CC recombination process to produce the polypeptide and antagonist against  
CC the polypeptide. The present invention also discloses the method of  
CC applying the polypeptide in treating various diseases, such as malignant  
CC tumors, inflammations, immunological diseases, haemopathy and HIV  
CC infection. The present sequence is the human dioxigenase 10.12.

XX Sequence 92 AA;

Query Match 52.1%; Score 405.5; DB 5; Length 92;  
Best Local Similarity 79.2%; Pred. No. 7.2e-33;  
Matches 80; Conservative 0; Mismatches 2; Indels 19; Gaps 1;

QY 45 GPEAGGLHSGMLEDGLPSNGVPRSTAPGGINPEKKTNCETQCPNPQSLSSGGLTKQKNG 104

DB 11 GPEAG-----NAFGGINPEKKTNCETQCPNPQSLSSGGLTKQKNG 51

QY 105 LQTTEAKRDAKRMPAKEVTINVTDISIQQMDRSRRITKNCVN 145

DB 52 LQTTEAKRDAKRMPAKEVTINVTDISIQQMDRSRRITKNCVN 92

RESULT 12

ADM46962

ID ADM46962 standard; protein; 73 AA.

AC ADM46962;

DT 03-JUN-2004 (first entry)

DE Brain and Acute Leukemia, Cytoplasmic alternate protein #4.

XX acute myelogenous leukemia; gene expression; BAALC;  
KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
KW Cytoplasmic; exon.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc-difference 41

FT /note= "encoded by GCS"

PN WO2003040347-A2.

PD 15-MAY-2003.

XX 12-NOV-2002; 2002WO-US036375.

PR 09-NOV-2001; 2001US-0348210P.

XX (OHIS ) UNIV OHIO STATE RES FOUND.

PI Tanner SM, De La Chapell A;

DR WPI; 2003-441564/41.

DR N-PSDB; ADM46954.

XX Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
PT in a patient comprises assaying for the overexpression of one or more  
PT BAALC transcripts in cells obtained from the patient.

PS Disclosure; SEQ ID NO 20; 78pp; English.

XX The invention relates to a method of characterizing acute myelogenous  
XX leukemia (AML) in a patient by assaying for the overexpression of one or  
XX more BAALC transcripts in cells obtained from the patient, where an  
XX overexpression indicates that the patient has an aggressive form of AML.  
XX The methods, kits and probes are useful for characterizing acute or  
XX chronic myelogenous leukemia, or prostate cancer. They are also useful  
XX for detecting BAALC overexpression. This sequence corresponds to a BAALC  
XX (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
XX spliced RNA consisting of exons 1, 6 and 8.

SQ Sequence 73 AA;

Query Match 39.7%; Score 309; DB 7; Length 73;  
 Best Local Similarity 96.6%; Pred. No. 2,8e-23;  
 Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDDAPPSAAAPDSGPEAGGLHSGMLE 58  
 DB 1 MCGGSRADAEIPRYESWTRETESTWLTYSDDAPPSAAAPDSGPEAGGLHSGCLE 58

RESULT 13

ADM46960  
 ID ADM46960 standard; protein; 54 AA.  
 XX  
 AC ADM46960;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Brain and Acute Leukemia, Cytoplasmic alternate protein #2.  
 XX  
 KW acute myelogenous leukemia; gene expression; BAALC;  
 KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 KW Cytoplasmic; exon.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 41 /note= "encoded by GCS"  
 FT  
 XX WO2003040347-A2.  
 PN  
 XX 15-MAY-2003.  
 PD  
 XX 12-NOV-2002; 2002WO-US036375.  
 PF  
 XX 09-NOV-2001; 2001US-0348210P.  
 FR  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 PA  
 XX Tanner SM, De La Chapell A;  
 PI  
 XX WPI; 2003-441564/41.  
 DR  
 XX N-PSDB; ADM46952.  
 XX  
 Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 PT in a patient comprises assaying for the overexpression of one or more  
 PT BAALC transcripts in cells obtained from the patient.  
 XX  
 Disclosure; SEQ ID NO 18; 78pp; English.

The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
 CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.

SQ Sequence 54 AA;

Query Match 38.4%; Score 299; DB 7; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-22;  
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDDAPPSAAAPDSGPEAGGLHSG 54  
 DB 1 MCGGSRADAEIPRYESWTRETESTWLTYSDDAPPSAAAPDSGPEAGGLHSG 54

RESULT 14

ADO48477  
 ID ADO48477 standard; protein; 54 AA.  
 XX  
 AC ADO48477;  
 XX  
 DT 12-AUG-2004 (first entry)  
 XX  
 DE Human PTH responsive gene protein exon 2 splice variant.  
 XX  
 KW PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
 KW transgenic animal; osteopathic; gene therapy; osteoporosis; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004044152-A2.  
 XX  
 DT 27-MAY-2004.  
 FD  
 XX 10-NOV-2003; 2003WO-US035655.  
 PF  
 XX 12-NOV-2002; 2002US-0425532P.  
 PR  
 XX (AMHP ) WYETH.  
 XX  
 PA Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;  
 PI  
 XX WPI; 2004-420299/39.  
 DR  
 XX N-PSDB; ADO48476.  
 DR  
 XX  
 New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
 PT preparing a composition for diagnosing, treating or preventing bone  
 PT related disorders, e.g., osteoporosis.  
 PT  
 XX  
 Claim 9; SEQ ID NO 6; 169pp; English.

The invention relates to a novel PTH responsive gene (PAIGB) fragment  
 CC encoding a polypeptide. The invention further comprises: a chimeric  
 CC construct comprising the isolated nucleic acid fragment operatively  
 CC linked to suitable regulatory sequences; a host cell transformed with the  
 CC chimeric construct; a vector comprising the nucleic acid fragment;  
 CC obtaining a nucleic acid fragment encoding the polypeptide; a method for  
 CC obtaining a polypeptide; detecting the presence of the nucleic acid  
 CC fragment; an antibody that specifically binds to one or more epitopes of  
 CC a PAIGB polypeptide; a composition for regulating bone-forming activity  
 CC in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
 CC ; an agent that alters the expression of PAIGB gene or polypeptide;  
 CC determining whether an agent alters the expression of PAIGB mRNA;  
 CC screening agents for effectiveness in altering expression of the nucleic  
 CC acid fragment; evaluating the efficacy of a treatment of a bone related  
 CC disorders; screening for agents useful for treating bone related  
 CC disorders in a subject; identifying polypeptides capable of binding to  
 CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
 CC related agent; a transgenic animal comprising the DNA; an animal model  
 CC for the study of bone density modulation comprising a first group of  
 CC animals composed of the transgenic animal and a second group of control  
 CC animals; studying bone mass determinants; studying the modulation of bone  
 CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
 CC for treating bone related disorders; identifying whether an agent which  
 CC has bone forming activity; and a stably transfected cell line comprising  
 CC two constructs, the first construct comprising a ligand binding domain  
 CC linked to a DNA binding domain which is linked to an activation domain  
 CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

XX Sequence 54 AA;

Job time : 101 secs

Query Match 38.4%; Score 299; DB 8; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-22;  
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCGGSRADAIERYESWTRETESTWLTYSDDAPPSSAAPDSGPEAGGLHSG 54  
 |||||  
 DB 1 MCGGSRADAIERYESWTRETESTWLTYSDDAPPSSAAPDSGPEAGGLHSG 54  
 |||||

RESULT 15  
 ADM46964  
 ID ADM46964 standard; protein; 80 AA.  
 XX  
 AC ADM46964;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Brain and Acute Leukemia, Cytoplasmic alternate protein #6.  
 XX  
 XX acute myelogenous leukemia; gene expression; BAALC;  
 XX chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 KW Cytoplasmic; exon.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 41  
 FT /note= "encoded by GCS"  
 XX  
 XX WO2003040347-A2.  
 XX  
 PD 15-MAY-2003.  
 XX  
 XX 12-NOV-2002; 2002WO-US036375.  
 XX  
 XX 09-NOV-2001; 2001US-0348210P.  
 XX  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 XX Tanner SM, De La Chapell A;  
 XX  
 XX WPI; 2003-441564/41.  
 DR N-PSDB; ADM46956.  
 XX  
 XX Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 FT in a patient comprises assaying for the overexpression of one or more  
 FT BAALC transcripts in cells obtained from the patient.  
 XX  
 PS Disclosure; SEQ ID NO 22; 78pp; English.  
 XX  
 CC The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
 CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.  
 XX  
 SQ Sequence 80 AA;

Query Match 37.7%; Score 293; DB 7; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-21;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCGGSRADAIERYESWTRETESTWLTYSDDAPPSSAAPDSGPEAGGLHSG 53  
 |||||  
 DB 1 MCGGSRADAIERYESWTRETESTWLTYSDDAPPSSAAPDSGPEAGGLHSG 53  
 |||||

Search completed: November 17, 2004, 15:22:24



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 15:22:38 ; Search time 81.3333 Seconds  
(without alignments)  
631.334 Million cell updates/sec

Title: US-10-705-716A-4

Perfect score: 778

Sequence: 1 MGCGRADALEPRYESWT.....VTDSIQQMDRSRRITKNCVN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	778	100.0	145	14	US-10-293-239-17	Sequence 17, Appl
2	778	100.0	145	14	US-10-177-390-30	Sequence 30, Appl
3	750.5	96.5	180	14	US-10-293-239-19	Sequence 19, Appl
4	569.5	73.2	149	14	US-10-293-239-21	Sequence 21, Appl
5	309	39.7	73	14	US-10-293-239-20	Sequence 20, Appl
6	299	38.4	54	14	US-10-293-239-18	Sequence 18, Appl
7	293	37.7	80	14	US-10-293-239-22	Sequence 22, Appl
8	140	18.0	25	14	US-10-293-239-37	Sequence 37, Appl
9	99	12.7	18	14	US-10-293-239-35	Sequence 35, Appl
10	92.5	11.9	307	16	US-10-437-963-181279	Sequence 3, Appl
11	92.5	11.9	1001	15	US-10-415-147-3	Sequence 14107, A
12	91	11.7	670	14	US-10-156-761-13447	Sequence 13447, A
13	90	11.6	219	14	US-10-156-761-13447	

14	88	11.3	298	10	US-09-910-600-26	Sequence 26, Appl
15	87.5	11.2	369	14	US-10-036-542-84	Sequence 84, Appl
16	87.5	11.2	369	15	US-10-188-832-149	Sequence 149, App
17	87.5	11.2	383	14	US-10-029-386-32058	Sequence 32058, A
18	87.5	11.2	572	17	US-10-425-115-340590	Sequence 340590,
19	87	11.2	136	15	US-10-424-599-179100	Sequence 179100,
20	86	11.1	182	16	US-10-767-701-37779	Sequence 37779, A
21	86	11.1	342	15	US-10-112-944-914	Sequence 914, App
22	86	11.1	668	16	US-10-437-963-134726	Sequence 134726,
23	85.5	11.0	496	15	US-10-425-114-71015	Sequence 71015, A
24	85	10.9	486	15	US-10-282-122A-62114	Sequence 62114, A
25	84	10.8	514	9	US-09-712-363-154	Sequence 154, App
26	83.5	10.7	332	16	US-10-437-963-168390	Sequence 168390,
27	83.5	10.7	558	14	US-10-156-761-12110	Sequence 12110, A
28	83	10.7	200	16	US-10-767-701-32315	Sequence 32315, A
29	83	10.7	373	16	US-10-437-963-147910	Sequence 147910,
30	82.5	10.6	497	16	US-10-437-963-157852	Sequence 157852,
31	82.5	10.6	850	15	US-10-424-599-242653	Sequence 242653,
32	82	10.5	216	16	US-10-767-701-57343	Sequence 57343, A
33	82	10.5	657	16	US-10-437-963-163549	Sequence 163549,
34	82	10.5	795	15	US-10-424-599-174901	Sequence 174901,
35	81.5	10.5	326	15	US-10-425-114-69350	Sequence 69350, A
36	81	10.4	147	16	US-10-767-701-48905	Sequence 48905, A
37	81	10.4	243	15	US-10-424-599-209325	Sequence 209325,
38	81	10.4	355	15	US-10-425-114-42733	Sequence 42733, A
39	81	10.4	2527	16	US-10-408-765A-2462	Sequence 2462, App
40	81	10.4	2715	16	US-10-408-765A-866	Sequence 866, App
41	80.5	10.3	213	15	US-10-424-599-190791	Sequence 190791,
42	80.5	10.3	337	16	US-10-437-963-107097	Sequence 107097,
43	80.5	10.3	342	9	US-09-269-390-4	Sequence 4, Appl
44	80.5	10.3	342	14	US-10-317-832-118	Sequence 118, App
45	80.5	10.3	342	17	US-10-733-878-118	Sequence 118, App

#### ALIGNMENTS

#### RESULT 1

US-10-293-239-17  
; Sequence 17, Application US/10293239  
; Publication No. US20030119043A1  
; GENERAL INFORMATION:  
; APPLICANT: Tanner, Stephan  
; APPLICANT: de la Chapelle, Albert  
; TITLE OF INVENTION: BAAC expression as a diagnostic marker for acute leukemia  
; FILE REFERENCE: 22727/04101  
; CURRENT APPLICATION NUMBER: US/10/293,239  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/348,210  
; PRIOR FILING DATE: 2001-11-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 145  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-293-239-17

Query Match 100.0%; Score 778; DB 14; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1.4e-66;  
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGCGRADALEPRYESWTRETETWTYDSDAPSAAPDSGPEAGGLHSGNLEGL	60
Db	1	MGCGRADALEPRYESWTRETETWTYDSDAPSAAPDSGPEAGGLHSGNLEGL	60
Qy	61	PSNGVPRSTAPGIPNPEKKNCTQCPNPSLSSGGLPTQKQGLQTTAKRDKRMFAK	120
Db	61	PSNGVPRSTAPGIPNPEKKNCTQCPNPSLSSGGLPTQKQGLQTTAKRDKRMFAK	120
Qy	121	EVTINVTDSIQQMDRSRRITKNCVN	145
Db	121	EVTINVTDSIQQMDRSRRITKNCVN	145



## RESULT 2

US-10-177-390-30  
; Sequence 30, Application US/10177390  
; Publication No. US20030143743A1  
; GENERAL INFORMATION:  
; APPLICANT: Schuler, Gerold  
; APPLICANT: N.V. Antwerp8 Innovatiecentrum  
; TITLE OF INVENTION: Improved Transfection of Eucaryotic Cells with Linear  
; FILE OF INVENTION: Polynucleotides by Electroporation  
; FILE REFERENCE: 021505wo/JH/ml  
; CURRENT APPLICATION NUMBER: US/10/177,390  
; CURRENT FILING DATE: 2002-06-20  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 30  
; LENGTH: 145  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-177-390-30

Query Match 100.0%; Score 778; DB 14; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1.4e-66;  
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYYTDSADPPSAAAPDSGPPAGGLHSGMLDGL 60  
Db 1 MCGGSRADAIEPRYYESWTRETESTWLTYYTDSADPPSAAAPDSGPPAGGLHSGMLDGL 60  
QY 61 PSNGVPRSTAPGGIPNPEKKTNCETQCPNPQSLSSGGLTQKQNGLOTTEAKRDAKMPAK 120  
Db 61 PSNGVPRSTAPGGIPNPEKKTNCETQCPNPQSLSSGGLTQKQNGLOTTEAKRDAKMPAK 120  
QY 121 EVTINVTDSIQMDRSRRITKNCVN 145  
Db 121 EVTINVTDSIQMDRSRRITKNCVN 145

## RESULT 3

US-10-293-239-19  
; Sequence 19, Application US/10293239  
; Publication No. US20030119043A1  
; GENERAL INFORMATION:  
; APPLICANT: Tanner, Stephan  
; APPLICANT: de la Chapelle, Albert  
; TITLE OF INVENTION: BAAC expression as a diagnostic marker for acute leukemia  
; FILE REFERENCE: 22727/04101  
; CURRENT APPLICATION NUMBER: US/10/293,239  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/348,210  
; PRIOR FILING DATE: 2001-11-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 180  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-293-239-19

Query Match 96.5%; Score 750.5; DB 14; Length 180;  
Best Local Similarity 80.6%; Pred. No. 8e-64;  
Matches 145; Conservative 0; Mismatches 0; Indels 35; Gaps 1;  
QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYYTDSADPPSAAAPDSGPPAGGLHSGMLDGL 53  
Db 1 MCGGSRADAIEPRYYESWTRETESTWLTYYTDSADPPSAAAPDSGPPAGGLHSGMLDGL 60  
QY 54 -----GMLDGLPSNGVPRSTAPGGIPNPEKKTNCET 85  
Db 61 KIKAPTDVSDEGLFSASKMAPLAVFSHGMLDGLPSNGVPRSTAPGGIPNPEKKTNCET 120  
QY 86 QCFNPQSLSSGGLTQKQNGLOTTEAKRDAKMPAKVITNTDSIQMDRSRRITKNCVN 145

Db 121 QCFNPQSLSSGGLTQKQNGLOTTEAKRDAKMPAKVITNTDSIQMDRSRRITKNCVN 180

## RESULT 4

US-10-293-239-21  
; Sequence 21, Application US/10293239  
; Publication No. US20030119043A1  
; GENERAL INFORMATION:  
; APPLICANT: Tanner, Stephan  
; APPLICANT: de la Chapelle, Albert  
; TITLE OF INVENTION: BAAC expression as a diagnostic marker for acute leukemia  
; FILE REFERENCE: 22727/04101  
; CURRENT APPLICATION NUMBER: US/10/293,239  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/348,210  
; PRIOR FILING DATE: 2001-11-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 21  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-293-239-21

Query Match 73.2%; Score 569.5; DB 14; Length 149;  
Best Local Similarity 75.7%; Pred. No. 1.4e-46;  
Matches 109; Conservative 0; Mismatches 0; Indels 35; Gaps 1;  
QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYYTDSADPPSAAAPDSGPPAGGLHSGMLDGL 53  
Db 1 MCGGSRADAIEPRYYESWTRETESTWLTYYTDSADPPSAAAPDSGPPAGGLHSGMLDGL 60  
QY 54 -----GMLDGLPSNGVPRSTAPGGIPNPEKKTNCET 85  
Db 61 KIKAPTDVSDEGLFSASKMAPLAVFSHGMLDGLPSNGVPRSTAPGGIPNPEKKTNCET 120  
QY 86 QCFNPQSLSSGGLTQKQNGLOTTE 109  
Db 121 QCFNPQSLSSGGLTQKQNGLOTTE 144

## RESULT 5

US-10-293-239-20  
; Sequence 20, Application US/10293239  
; Publication No. US20030119043A1  
; GENERAL INFORMATION:  
; APPLICANT: Tanner, Stephan  
; APPLICANT: de la Chapelle, Albert  
; TITLE OF INVENTION: BAAC expression as a diagnostic marker for acute leukemia  
; FILE REFERENCE: 22727/04101  
; CURRENT APPLICATION NUMBER: US/10/293,239  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/348,210  
; PRIOR FILING DATE: 2001-11-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-293-239-20

Query Match 39.7%; Score 309; DB 14; Length 73;  
Best Local Similarity 96.6%; Pred. No. 5.5e-22;  
Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYYTDSADPPSAAAPDSGPPAGGLHSGMLDGL 58  
Db 1 MCGGSRADAIEPRYYESWTRETESTWLTYYTDSADPPSAAAPDSGPPAGGLHSGMLDGL 58

## RESULT 6

```
US-10-293-239-18
; Sequence 18, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-18
Query Match      38.4%; Score 299; DB 14; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.4e-21;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCGGSRADAIEPRYYESWTRETESTLWTTDSDAPPSAAAPDSGPEAGGLHSG 54
Db 1 MCGGSRADAIEPRYYESWTRETESTLWTTDSDAPPSAAAPDSGPEAGGLHSG 54

RESULT 7
US-10-293-239-22
; Sequence 22, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-22
Query Match      37.7%; Score 291; DB 14; Length 80;
Best Local Similarity 100.0%; Pred. No. 2.1e-20;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCGGSRADAIEPRYYESWTRETESTLWTTDSDAPPSAAAPDSGPEAGGLHS 53
Db 1 MCGGSRADAIEPRYYESWTRETESTLWTTDSDAPPSAAAPDSGPEAGGLHS 53

RESULT 8
US-10-293-239-37
; Sequence 37, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-37
Query Match      18.0%; Score 140; DB 14; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.3e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 RADAIEPRYYESWTRETESTLWTT 31
Db 1 RADAIEPRYYESWTRETESTLWTT 25

RESULT 9
US-10-293-239-35
; Sequence 35, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-35
Query Match      12.7%; Score 99; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DAIEPRYYESWTRETEST 26
Db 1 DAIEPRYYESWTRETEST 18

RESULT 10
US-10-437-963-181279
; Sequence 181279, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kowalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 181279
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(307)
; OTHER INFORMATION: unsure at all Xaa locations
```

```

; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_78569C.1.1.psp
US-10-437-963-181279

Query Match      11.9%; Score 92.5; DB 16; Length 307;
Best Local Similarity 27.8%; Pred. No. 1.9;
Matches 37; Conservative 18; Mismatches 39; Indels 39; Gaps 8;

QY 33 SDAPPSAAAPDSGPEA-GGLHSGMLEDG-----LPSNGVPRSTAGGTPNP 77
Db 87 SGRPPPPRPPGAGADAVAGVHTSSLTAGTTSPQTTLRANSPLFS--LPRASTPP--PSP 142
QY 78 EKKTNCETCPNPQSLSSGP-----LTKQKQGLQTTAKRDA-----KMPAKEV 122
Db 143 PRLT--TTPTPTPSSSTPRLSLSRLRLLTTPPPRCILTSHARRPPLQTHDRRPLQSM 199
QY 123 TINVDSIQOMDR 135
Db 200 TVTXPDT--RLDR 210

RESULT 11
US-10-415-147-3
; Sequence 3, Application US/10415147
; Publication No. US20040043399A1
; GENERAL INFORMATION:
; APPLICANT: Bayer AG
; TITLE OF INVENTION: REGULATION OF HUMAN INOSITOL POLYPHOSPHATE 5-PHOSPHATASE
; FILE REFERENCE: L10235Foreign Countries
; CURRENT APPLICATION NUMBER: US/10/415,147
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: US 60/243,745
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: US 60/257,302
; PRIOR FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: US 60/314,660
; PRIOR FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-415-147-3

Query Match      11.9%; Score 92.5; DB 15; Length 1001;
Best Local Similarity 25.8%; Pred. No. 8;
Matches 32; Conservative 15; Mismatches 48; Indels 29; Gaps 4;

QY 12 EPRYVESWTRETSTWLTYYTDSAPP-----SAAAPDSGPEAGLHSGMLEDGLPS 62
Db 275 DPLSPSFRARPEARHSPEDVLPFPPTQLPLDVSSGLPESGRSPGLLSPTRPFGIPS 334
QY 63 NGVPRSTAPGGIPNPEK-----KTNC-----ETQCPNPQSLSSGGLTKQKGLQ 106
Db 335 N---QTVPPPLPKPRSPRSRSPNRSPCVPPAPEVALPRPVTTGAGKCKPCPSNLQ 390
QY 107 TTEA 110
Db 391 TQES 394

RESULT 12
US-10-156-761-14107
; Sequence 14107, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13447
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13447

Query Match      11.6%; Score 90; DB 14; Length 219;
Best Local Similarity 33.9%; Pred. No. 2.1;
Matches 20; Conservative 8; Mismatches 31; Indels 0; Gaps 0;

QY 6 SRADAIEPRYVESWTRETSTWLTYYTDSAPPAAAPDSGPEAGLHSGMLEDGLPSNG 64
Db 43 ARAGVGKQTIYRWWSKAEVLLEAFTLSAQAAEAAARPGPEGGGGQGGGQENGIPDTG 101

RESULT 14
US-09-910-600-26
; Sequence 26, Application US/09910600
; Publication No. US20030036631A1

```

Db 96 VVASPLDQDEGSSSQKEESPSTLQVLPSSELSRSEIDEKVTDLVQ 143

Search completed: November 17, 2004, 15:48:53  
Job time : 82.3333 secs

```

; GENERAL INFORMATION:
; APPLICANT: Longphre, Malinda
; APPLICANT: Chang, Han
; APPLICANT: Whitney, Gena
; TITLE OF INVENTION: NOVEL SIGLECS AND USES THEREOF
; FILE REFERENCE: D0003NP
; CURRENT APPLICATION NUMBER: US/09/910,600
; PRIORITY FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/220,139
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 26
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: L3cyto-Y641
; OTHER INFORMATION: alone
US-09-910-600-26

Query Match          11.3%; Score 88; DB 10; Length 298;
Best Local Similarity 27.5%; Pred. No. 4.8;
Matches 33; Conservative 18; Mismatches 43; Indels 26; Gaps 7;

Qy      3  CGGSRADAI-----EPYYESWTEYESTW-LTYTSDAPPSA-----AAPDS---GP 46
Db      178  CFKRIEALPQIDKYLSKSYIAMPDQ---GQNTFGGDDHPFKSDLVPRGSPMSINVVP 234

Qy      47  EAGLHSGMLEDGLPENGVPSTAPGGIPNPEKKTNCETQ-----CPNPQSLSSGGLPTQK 101
Db      235  TAGPLAQENKQKATPNS--PRTPLPPGAPSPESKKNOKKQYOLPSPFKPSGSTOAPPSOE 292

```

RESULT 15  
 US-10-036-542-84  
 ; Sequence 84, Application US/10036542  
 ; Publication No. US20030083481A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Birse et al.  
 ; TITLE OF INVENTION: 25 Human Prostate and Prostate Cancer Associated Proteins  
 ; FILE REFERENCE: PA002PI  
 ; CURRENT APPLICATION NUMBER: US/10/036,542  
 ; CURRENT FILING DATE: 2002-01-07  
 ; PRIOR APPLICATION NUMBER: PCT/US00/19666  
 ; PRIOR FILING DATE: 2000-07-20  
 ; PRIOR APPLICATION NUMBER: 60/144,972  
 ; PRIOR FILING DATE: 1999-07-21  
 ; PRIOR APPLICATION NUMBER: 60/148,681  
 ; PRIOR FILING DATE: 1999-08-13  
 ; PRIOR APPLICATION NUMBER: 60/149,173  
 ; PRIOR FILING DATE: 1999-08-17  
 ; PRIOR APPLICATION NUMBER: 60/158,004  
 ; PRIOR FILING DATE: 1999-10-06  
 ; PRIOR APPLICATION NUMBER: 60/194,689  
 ; PRIOR FILING DATE: 2000-04-05  
 ; NUMBER OF SEQ ID NOS: 157  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 84  
 ; LENGTH: 369  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-036-542-84

```

Query Match      11.2%; Score 87.5; DB 14; Length 369;
Best Local Similarity 25.3%; Pred. No. 7;
Matches 28; Conservative 19; Mismatches 48; Indels 13; Gaps 4;

Qy 33 SDAPPAAADSGPEAGGLHSGMLDGLSPGVPRS--TAPGGIPNPKKTKNCETQCENPQ 91
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 40 SSTSTSSSPSPSPSSSSSSSSSSSSSSCYLIPSS--TPPEVSADDDETPNPQ--SAQIACSPPS 95
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 92 SLSSGGSL-----TQKNGIQLTTEAKRDAKRMPEKVTINVTDSIQ 131

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 15:02:57 ; Search time 24.3333 Seconds  
(without alignments)  
395.183 Million cell updates/sec

Title: US-10-705-716A-4

Perfect score: 778

Sequence: 1 MCGGSRADAIEPRYESWT.....VTDSIQMDRSRRITKNVCN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87.5	11.2	369	2	US-08-773-870-4
2	85.5	11.0	316	4	US-09-248-796A-16753
3	81	10.4	208	4	US-09-252-991A-25785
4	80	10.3	521	4	US-09-086-663A-81
5	80	10.3	528	4	US-09-086-663A-82
6	80	10.3	548	4	US-09-086-663A-71
7	80	10.3	596	4	US-09-086-663A-2
8	80	10.3	596	4	US-09-086-663A-80
9	78.5	10.1	330	4	US-09-252-991A-21479
10	78.5	10.1	462	4	US-09-976-594-427
11	78	10.0	230	3	US-09-248-335-44
12	78	10.0	264	1	US-08-562-311-4
13	78	10.0	312	4	US-09-252-991A-28271
14	78	10.0	687	4	US-09-248-796A-23026
15	78	10.0	2016	3	US-09-634-920-4
16	78	10.0	2016	4	US-09-514-907A-2
17	78	10.0	2016	4	US-09-896-994-2
18	78	10.0	2016	4	US-09-840-125-4
19	77.5	10.0	286	4	US-09-071-035-176
20	77.5	10.0	305	4	US-09-071-035-174
21	77	9.9	160	4	US-09-621-976-4200
22	77	9.9	217	4	US-09-252-991A-21052
23	77	9.9	437	4	US-09-248-796A-18684
24	77	9.9	441	4	US-09-248-796A-20171
25	77	9.9	707	4	US-09-919-039-278
26	77	9.9	707	4	US-09-538-092-993
27	77	9.9	1088	3	US-09-130-242-2

28 77 9.9 1088 4 US-09-583-610D-2 Sequence 2, Appli  
29 76.5 9.8 254 4 US-09-252-991A-20551 Sequence 20551, A  
30 76.5 9.8 277 4 US-09-252-991A-17311 Sequence 17311, A  
31 76.5 9.8 381 4 US-09-252-991A-29050 Sequence 29050, A  
32 76 9.8 512 4 US-09-496-320-11 Sequence 11, Appli  
33 76 9.8 1873 1 US-08-435-675B-4 Sequence 4, Appli  
34 76 9.8 1873 1 US-08-336-257A-7 Sequence 7, Appli  
35 75.5 9.7 109 4 US-09-252-991A-25380 Sequence 25380, A  
36 75.5 9.7 402 4 US-09-252-991A-29857 Sequence 29857, A  
37 75.5 9.7 641 4 US-09-252-991A-26994 Sequence 26994, A  
38 75.5 9.7 655 1 US-08-148-910-12 Sequence 12, Appli  
39 75.5 9.7 655 1 US-08-448-937A-12 Sequence 12, Appli  
40 75.5 9.7 941 4 US-07-757-022B-14 Sequence 14, Appli  
41 75.5 9.7 1022 4 US-07-757-022B-84 Sequence 84, Appli  
42 75.5 9.7 1038 4 US-07-757-022B-74 Sequence 74, Appli  
43 75.5 9.7 1049 4 US-07-757-022B-58 Sequence 58, Appli  
44 75.5 9.7 1140 4 US-07-757-022B-104 Sequence 104, App  
45 75.5 9.7 1270 4 US-07-757-022B-44 Sequence 44, Appli

#### ALIGNMENTS

RESULT 1  
US-08-773-870-4  
; Sequence 4, Application US/08773870  
; Patent No. 5912143  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN MAGE-LIKE PROTEIN  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/08/773,870  
; APPLICATION NUMBER: US/08/773,870  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0179 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 369 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 533511  
US-08-773-870-4

Query Match 11.2%; Score 87.5; DB 2; Length 369;  
Best Local Similarity 25.9%; Pred. NO. 0.34;  
Matches 28; Conservative 19; Mismatches 48; Indels 13; Gaps 4;

QY 33 SDAPPAAPDSGPEAGLHSGMLEDGLPSNGVPRS-TAPGGIPNPEKTKNCETQCPNPQ 91  
Db 40 SSTSTSSPSPSPSSSSSSSCYPLIPS--TPREVSADDETNPQPQ--SAQIACSSPS 95  
QY 92 SLSSGPI-----TQKQNGLOTTAKRDAKMPAKEVTINVTDSIQ 131  
Db 96 VVASLPDQSDGSSQKERSPTLQVLPSLSRSEIDKVTDLVQ 143

RESULT 2  
US-09-248-796A-16753  
; Sequence 16753, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 16753  
; LENGTH: 316  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-16753

Query Match 11.0%; Score 85.5; DB 4; Length 316;  
Best Local Similarity 31.5%; Pred. No. 0.45;  
Matches 29; Conservative 12; Mismatches 38; Indels 13; Gaps 5;

QY 63 NGVPRS--TAPGGIPNPEKTKNCE-TQCPNPQSL-SSGPLTQKONGLOTTAKRDA--KR 116  
Db 76 NFTPQSTDTPAATAVAKSNPKTNAEPAKIPNEKLLKTESPLSQKONGATTTKKESDVLLT 135  
QY 117 MPAKEVTINVTDSIQMDR-----SRRITK 141  
Db 136 KSTSTTVSNNSVLQYTSELSEIPIGVERITK 167

RESULT 3  
US-09-252-991A-25785  
; Sequence 25785, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25785  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25785

Query Match 10.4%; Score 81; DB 4; Length 208;  
Best Local Similarity 23.0%; Pred. No. 0.77;  
Matches 37; Conservative 18; Mismatches 70; Indels 36; Gaps 4;

QY 3 CGGSR---ADATPRYIESWTETESTWLTYTDSAPPSAAAPDSGPEAGLHSGMLED 58

Db 28 CGGSASIRPARAARPSPKPASSPTTAMTGSTRMTRPSTSAKARPTTTCGASASAPATTP 87  
QY 59 GL-----PSNGVPRSTAPGGIPNPEKTKNCETQ-----CPN 89  
Db 88 GTTGPTTTPAPTQKSAAGCAMPDCHACRRRTARPPPTPPVAPWISSSTRASVPGACTC 147  
QY 90 POSLSSGPLTQKONGLOTTAKRDAK---RMPAKEVTINVT 127  
Db 148 PTSRTGTSSRRRTTCTAPTCTCSRRYARRPAKATTRCT 188

RESULT 4  
US-09-086-663A-81  
; Sequence 81, Application US/09086663A  
; Patent No. 6518063  
; GENERAL INFORMATION:  
; APPLICANT: DUCY, PATRICIA  
; TITLE OF INVENTION: OSF2/CBFAL COMPOSITIONS AND METHODS OF USE  
; FILE REFERENCE: UTSC:525  
; CURRENT APPLICATION NUMBER: US/09/086,663A  
; CURRENT FILING DATE: 1998-05-29  
; PRIOR APPLICATION NUMBER: 60/080,189  
; PRIOR FILING DATE: 1998-03-24  
; PRIOR APPLICATION NUMBER: 60/048,430  
; PRIOR FILING DATE: 1997-05-29  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 81  
; LENGTH: 521  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-086-663A-81

Query Match 10.3%; Score 80; DB 4; Length 521;  
Best Local Similarity 29.4%; Pred. No. 3.6;  
Matches 30; Conservative 7; Mismatches 57; Indels 8; Gaps 2;

QY 24 ESTWLTYTSDAPPSAA-----APDSGPEAGLHSGMLEDGLPSNGVPRSTAPGGIPN 76  
Db 195 KSFTLITVTNPPQVATYHRAIKVTVDGPREPRHROKLLDSSKPSLFSRDLSDLGRIH 254  
QY 77 PEKTKNCETQCPNPQSLSGPLTQKONGLOTTAKRDAKMP 118  
Db 255 PSRMVGVPPQNP RP-SLNSAPSPFNPGOSQITDPRQAQSSP 295

RESULT 5  
US-09-086-663A-82  
; Sequence 82, Application US/09086663A  
; Patent No. 6518063  
; GENERAL INFORMATION:  
; APPLICANT: DUCY, PATRICIA  
; TITLE OF INVENTION: OSF2/CBFAL COMPOSITIONS AND METHODS OF USE  
; FILE REFERENCE: UTSC:525  
; CURRENT APPLICATION NUMBER: US/09/086,663A  
; CURRENT FILING DATE: 1998-05-29  
; PRIOR APPLICATION NUMBER: 60/080,189  
; PRIOR FILING DATE: 1998-03-24  
; PRIOR APPLICATION NUMBER: 60/048,430  
; PRIOR FILING DATE: 1997-05-29  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 82  
; LENGTH: 528  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-086-663A-82

Query Match 10.3%; Score 80; DB 4; Length 528;  
Best Local Similarity 29.4%; Pred. No. 3.6;  
Matches 30; Conservative 7; Mismatches 57; Indels 8; Gaps 2;



ORGANISM: *Pseudomonas aeruginosa*

US-09-252-991A-21479

Query Match 10.1%; Score 78.5; DB 4; Length 330;  
 Best Local Similarity 26.2%; Pred. No. 2.7; Indels 33; Gaps 6;  
 Matches 34; Conservative 13; Mismatches 50; Indels 33; Gaps 6;  
 QY 2 GCGSRAADALE-----PRYYESTRETETSTWLTYSDDAPPSA-----39  
 Db 26 GCGSDRTPQLRKAGYAVEQPRLYGKVAEASATGLSITVPLPSPRNLCGAIVT 85  
 QY 40 -AAPDSGPEAGLHSGMLEDGLPSNGVPRSTAPCGINPEKKN--CETQCPNQSLSG 96  
 Db 86 DATPGLLPGGGFRHRRRDPSTHLPRS--PGG---KDRLLHLRRQAPAP---AAG 137  
 QY 97 PLTQKQNGLO 106  
 Db 138 PARRQDPGLQ 147

RESULT 10

US-09-976-594-427  
 ; Sequence 427, Application US/09976594  
 ; Patent No. 6673549  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Furness, Michael  
 ; APPLICANT: Buchbinder, Jenny  
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
 ; FILE REFERENCE: PA-0041 US  
 ; CURRENT APPLICATION NUMBER: US/09/976,594  
 ; CURRENT FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: 60/240,409  
 ; PRIOR FILING DATE: 2000-10-12  
 ; NUMBER OF SEQ ID NOS: 1143  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 427  
 ; LENGTH: 462  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. 6673549 2716815CD1  
 US-09-976-594-427

Query Match 10.1%; Score 78.5; DB 4; Length 462;  
 Best Local Similarity 28.6%; Pred. No. 4.4; Length 462;  
 Matches 30; Conservative 16; Mismatches 32; Indels 27; Gaps 7;  
 QY 16 YESWTRETSTWLTYSDDAPPSAAAPDSGPEAGLHSGMLEDGLPSNGVPRSTAPCGIP 75  
 Db 298 YRRMSAEVTS--TYSDEDRPKK--VPPREFLS-----PSNS--RTSPKSLP 339  
 QY 76 NPEKKTNCETQ--CPNPSQSGPLTQKQNGLOTTTEAKRDARMP 118  
 Db 340 SYLNGWMPPTQSPAPDKYSSKAL-QRQN-----SEGSASKVP 377

RESULT 11

US-09-248-335-44  
 ; Sequence 44, Application US/09248335  
 ; Patent No. 6096504  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCGONIGLE, BRIAN  
 ; APPLICANT: O'KEEF, DANIEL  
 ; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES  
 ; FILE REFERENCE: CL-1128-A  
 ; CURRENT APPLICATION NUMBER: US/09/248,335  
 ; CURRENT FILING DATE: 1999-02-10  
 ; EARLIER APPLICATION NUMBER: 08/924,759  
 ; EARLIER FILING DATE: 1997-September-05  
 ; NUMBER OF SEQ ID NOS: 74  
 ; SOFTWARE: Microsoft Word Version 7.0A  
 ; SEQ ID NO 44  
 ; LENGTH: 230

; TYPE: PRT  
 ; ORGANISM: maize  
 US-09-248-335-44  
 Query Match 10.0%; Score 78; DB 3; Length 230;  
 Best Local Similarity 26.7%; Pred. No. 1.9;  
 Matches 36; Conservative 9; Mismatches 38; Indels 52; Gaps 7;  
 QY 8 ADAIEPRYYESWTR-----ETESTWL-----TVTDSADAPPSA-----39  
 Db 102 ADVIDKKIYDQRLWKFEGEAREQAKDLVEVLETWGRSSPTLSLSAAAPSAWTLWC 161  
 QY 40 -AAPDSGPEAGLHSGMLEDGLPSNGVPRSTAPCGINPEKKN-----CET--OCPN 89  
 Db 162 PSRFGSSPTRSW-----AGSASRSTAPGSGWPGPRAAGSGRAWPRCPTLPCSS 210  
 QY 90 PQS-----LSSQPLTQK 101  
 Db 211 SSSSSRASSGSDRK 225

RESULT 12

US-08-562-311-4  
 ; Sequence 4, Application US/08562311  
 ; Patent No. 5801002  
 ; GENERAL INFORMATION:  
 ; APPLICANT: RAZ, AVRAHAM  
 ; TITLE OF INVENTION: A METHOD OF DETERMINING THE PROBABILITY  
 ; TITLE OF INVENTION: OF METASTASIS IN A CELL SAMPLE  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dykema Gossett  
 ; STREET: STE 505 N. Woodward  
 ; CITY: Bloomfield Hills  
 ; STATE: MI  
 ; COUNTRY: U.S.  
 ; ZIP: 48304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/562,311  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/188,225  
 ; FILING DATE:  
 ; APPLICATION NUMBER: US 07/681,242  
 ; FILING DATE: 04-APR-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/294,249  
 ; FILING DATE: 01-JUN-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: KELLY, ROBERT L.  
 ; REGISTRATION NUMBER: 31,843  
 ; REFERENCE/DOCKET NUMBER: 61,686-  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 810-540-0849  
 ; TELEFAX: 810-540-0763  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 264 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-562-311-4

Query Match 10.0%; Score 78; DB 1; Length 264;  
 Best Local Similarity 33.8%; Pred. No. 2.3;  
 Matches 26; Conservative 3; Mismatches 32; Indels 16; Gaps 4;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 15:00:21 ; Search time 19.6667 Seconds  
(without alignments)  
709.395 Million cell updates/sec

Title: US-10-705-716A-4

Perfect score: 778

Sequence: 1 MCGGSRADALEPRYESWT.....VTDSIQMDRSRRITKNCVN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_79:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89.5	11.5	1122	2 T47424	hypothetical prote
2	88.5	11.4	690	2 H86464	hypothetical prote
3	87.5	11.2	369	2 I38659	melanoma antigen M
4	87.5	11.2	860	2 A36717	unknown protein, 4
5	84.5	10.9	733	2 A45301	microtubule-associ
6	84	10.8	416	1 W2WLDP	E2 protein - deer
7	84	10.8	514	2 H70699	probable ppp prote
8	84	10.8	672	2 I40333	tracheal colonizat
9	83.5	10.7	327	2 A96619	crta protein - rho
10	83.5	10.7	1115	1 IJMSNL	neural cell adhesi
11	82.5	10.6	1366	2 T35985	probable large pro
12	81.5	10.5	499	2 S22571	integrase-like pro
13	81.5	10.5	2793	2 B90784	hypothetical prote
14	81.5	10.5	2806	2 D85644	hypothetical prote
15	80.5	10.3	263	2 A85359	translation initia
16	80.5	10.3	327	2 T50744	spheroidene monoox
17	80	10.3	476	2 T32728	hypothetical prote
18	80	10.3	513	2 A48233	polyomavirus enhan
19	80	10.3	908	2 A33280	sarcaluminen precu
20	79.5	10.2	1008	2 T41244	SEC14 protein homo
21	79.5	10.2	1063	2 T03743	bifocal protein -
22	79.5	10.2	1494	2 T43355	protein-tyrosine-p
23	79	10.2	444	2 A25112	hypothetical prote
24	79	10.2	1006	2 T42731	atrophin-1 related
25	79	10.2	2157	2 S71461	proline-rich prote
26	79	10.2	2187	2 T30826	nascent polypeptid
27	78	10.0	197	2 T45983	probable periplasm
28	78	10.0	264	2 A45983	lactose-binding le
29	78	10.0	396	2 T35659	probable transmem

30	78	10.0	420	1 Q0BE44	BDLF2 protein - hu
31	78	10.0	1187	1 JC4155	protein-tyrosine-p
32	78	10.0	1216	2 F88473	protein F40H6.5 [i
33	78	10.0	2016	2 A38195	sodium channel pro
34	77.5	10.0	239	2 A27225	calmodulin-binding
35	77.5	10.0	464	2 A47655	spliceosome-associ
36	77.5	10.0	906	2 T00039	hypothetical prote
37	77.5	10.0	2109	1 I50421	aggreacan precursor
38	77.5	10.0	3511	2 A59295	unconventional myo
39	77	9.9	297	2 JQ1205	attachment protein
40	77	9.9	377	2 A48018	mucin 7 precursor,
41	77	9.9	707	2 A46302	PTB-associated spl
42	77	9.9	1100	2 H71616	SERA antigen/papai
43	76.5	9.8	264	2 C83632	hypothetical prote
44	76.5	9.8	421	2 A60058	neural cell adhesi
45	76.5	9.8	670	2 T49510	fibroin-3 related

#### ALIGNMENTS

##### RESULT 1

T47424

hypothetical protein T22K7.20 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004

C/Accession: T47424

R/Kieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.

submitted to the Protein Sequence Database, April 2000

A/Reference number: Z24459

A/Accession: T47424

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1122 <RIE>

A/Cross-references: UNIPROT:Q9M291; EMBL:AL138641

A/Experimental source: cultivar Columbia; BAC clone T22K7

C/Genetics:

A/Map position: 3

A/Introns: 303/3; 363/3; 388/3; 421/3; 459/1; 501/3; 552/3; 579/3; 605/3; 636/3; 655/1; 6

A/Note: T22K7.20

Query Match 11.5%; Score 89.5; DB 2; Length 1122;

Best Local Similarity 28.4%; Pred. No. 10;

Matches 23; Conservative 10; Mismatches 31; Indels 17; Gaps 3;

QY 36 PPSAAAP-----DSGPEAGGLHSGMLEDLGPNGVPRS-----TAPGGIPNPEKKT 81

DB 182 PPSGMPGGPLSNGPPSPSGMHGHLNSGPPSGMPGGLSNGPPSPSGMPGAFPRGSQFT 241

QY 82 NCETQCCNP---QSLSGGLT 99

DB 242 SGPMAPPFPYQPPNAGPFT 262

##### RESULT 2

H86464

hypothetical protein F12G12.9 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C/Accession: H86464

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

C/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, H.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: H86464

A/Status: preliminary

unknown protein, 45065-49536 [imported] - Arabidopsis thaliana  
A96717  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: A96717  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizars, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
A.

Db 270 EE 271

## RESULT 6

W2WLPD

C:Species: deer papillomavirus

C:Species: host Odocoileus virginianus (American white-tailed deer)

C:Date: 28-May-1986 #sequence\_revision 28-May-1986 #text\_change 24-Feb-1994

C:Accession: A03673

R:Groff, D.E.; Lancaster, W.D.

J. Virol. 56, 85-91, 1985

A:Title: Molecular cloning and nucleotide sequence of deer papillomavirus.

A:Reference number: A33013; MUID:85293253; PMID:2993669

A:Accession: A03673

A:Molecule type: DNA

A:Residues: 1-416 &lt;GRO&gt;

A:Superfamily: papillomavirus E2 protein

C:Keywords: early protein

Query Match 10.8%; Score 84; DB 1; Length 416;

Best Local Similarity 31.7%; Pred. No. 9.6;

Matches 38; Conservative 10; Mismatches 44; Indels 28; Gaps 7;

QY 4 GGSRADAI-----EPRIYESTRE-----TESTWLTYSADAP-PSAAAPDSGPEA-- 48

Db 152 GGADADGLFYTMSTGTRVYELFERDAARYSTGTW-IVRNDRTYHSHSAPSHRETIE 210

QY 49 GGLHSGMLEDLPSNGVPRS---TAPGG-----IPNEKTKTCTQCPNPSLSSG 96

Db 211 GLWNSGGRGRPTNSPDRAVLHTPPGNTVHGPRACENRGRSINRPTTPTQSPRS 270

## RESULT 7

H70699

Probable ppp protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004

C:Accession: H70699

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Rajandream, M.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: H70699

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-514 &lt;COL&gt;

A:Cross-references: UNIPROT:P71588; GB:Z80233; GB:AL123456; NID:g3261645; PIDN:CAB02438.

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: ppp

F:35-243/Domain: conserved hypothetical protein yloO homology &lt;YLOO&gt;

Query Match 10.8%; Score 84; DB 2; Length 514;

Best Local Similarity 25.9%; Pred. No. 12;

Matches 36; Conservative 10; Mismatches 49; Indels 44; Gaps 6;

QY 1 MGCGSRADAIPRIYESTRETESTWLTYSADAP-----PSAAAPDSGPEAGGLH 52

Db 357 MGC-----LSPR-----NELSQISYSGSGGLDCHLMKLEDLRPPERAQVRAGLP 401

QY 53 SCMLEDG-----LPSNGVPRSTAPGGINPEKTKTCTQCPNPSLSSGPTQ 100

Db 402 AGTLDDAIQGLRELAANLLPCCAPRATSPGPRAP--PTTSETTEPNVTSSPASP--- 456

QY 101 KQNGLOTTEAKRKDMA 119

Db 457 -----SPTTSAPAPTGTTPA 471

## RESULT 8

I40333

tracheal colonization factor A precursor - Bordetella pertussis

N:Alternate names: tcfA protein

C:Species: Bordetella pertussis

C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004

C:Accession: S77633; I40333

R:Finn, T.M.; Stevens, L.A.

Mol. Microbiol. 16, 625-634, 1995

A:Title: Tracheal colonization factor: a Bordetella pertussis secreted virulence determini

A:Reference number: I40333; MUID:96065692; PMID:7476158

A:Accession: S77633

A:Molecule type: DNA

A:Residues: 1-672 &lt;FIN&gt;

A:Cross-references: UNIPROT:Q45343; EMBL:U16754; NID:g984282; PIDN:AAC43453.1; PID:g9842

A:Experimental source: strain 18323

C:Genetics:

A:Gene: tcfA

F:1-39/Domain: signal sequence #status predicted &lt;SIG&gt;

F:40-672/Product: tracheal colonization factor A #status predicted &lt;MAT&gt;

Query Match 10.8%; Score 84; DB 2; Length 672;

Best Local Similarity 31.2%; Pred. No. 17;

Matches 35; Conservative 11; Mismatches 44; Indels 22; Gaps 6;

QY 30 YTSDSDAPP-----SAAAPDSGPEAGLHSGMLEDLPSNGVPRSTA--PGGI----PNPE 78

Db 103 HKDNPSPVVGVGPMAGSSGHNPGVGGTHENGPGIGKVGGSAPGPGGLGRNDENSE 162

QY 79 KTKTCTQCPNP---QSLSSGELTQKQNGLOTTE-----AKRDAKRMFAKE 121

Db 163 SSLNPGTLGSPGPDSTGTSGPDAGMASGAGSTSPGSGGAGKDA--MPPSE 212

## RESULT 9

S49619

crfA protein - Rhodobacter sphaeroides

C:Species: Rhodobacter sphaeroides

C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004

C:Accession: S49619

R:Lang, H.P.; Cogdell, R.J.; Takaichi, S.; Hunter, C.N.

submitted to the EMBL Data Library, November 1994

A:Description: The complete DNA sequence, specific TNS insertion map and gene assignment

A:Reference number: S49619

A:Accession: S49619

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-327 &lt;LAN&gt;

A:Cross-references: UNIPROT:Q54790; EMBL:X82458; NID:g575405; PID:g575406

C:Genetics:

A:Gene: crfA

C:Superfamily: spheroidene monooxygenase

Query Match 10.7%; Score 83.5; DB 2; Length 327;

Best Local Similarity 23.9%; Pred. No. 8;

Matches 33; Conservative 15; Mismatches 49; Indels 41; Gaps 5;

QY 2 CGGSRADAIPRIYESTRE-----TESTW-----LTYTSDAPPSAAAP 42

Db 190 GAGGPHGEAIKAVRAENWNFKELYARFOILGTICKWEGKDPVGEALTAAPSEAPKPAPAP 249

QY 43 DSGPEAGLHSGMLEDLPSNGVPRSTAPGGINPEKTKTCTQCPN---PQSLSSGP 97

Db 250 AAAQPA-----PAAEAPKAPAPVAEKPALAVEMPKPAPKPVVEAP 292

QY 98 LTKQNGLOTTEAKRKDAK 115

Db 293 KPFRSGLEADAAGREAE 310

## RESULT 10

IJMSNL

neural cell adhesion molecule 1 precursor, long domain splice form - mouse





```
Matches 40; Conservative 20; Mismatches 67; Indels 31; Gaps 9;
QY 13 PRYVES-----WTR--ETESTWL-----TYTSDAPPSAAPDSG--PE-----AGG 50
Db 35 PNAETDSDLGAGMVRVQDTSGTYWHLPTGTOWEPGRASPSQNSPQESQJTWTF 94
QY 51 LHSGMLEDLGSPNGVPRSTAPQ--GIPNPEKKTNCETQCP--NPQSLSSGPLTKQKNGIQT 107
Db 95 AHQSGFERGEWKPDESEAPWELGLXDPBEGT-----LPFSAQSLSPFPVQEEENLPQ 149
QY 108 TEAKRDAKMPAKEV-TINVTDSIQMDRSRRITKNCV 144
Db 150 RNANPGIKCFVRSGLGWEMTEELAPGRSSVAVNNCI 187
RESULT 13
B90784
Hypothetical protein Ecs1242 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: B90784
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90784
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2793 <HAY>
A:Cross-references: UNIPROT:Q8X2Q2; GB:BA000007; PIDN:BA834665.1; PID:g13360702; GSPDB:G
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: Ecs1242
Query Match 10.5%; Score 81.5; DB 2; Length 2793;
Best Local Similarity 27.3%; Pred. No. 1.4e+02;
Matches 30; Conservative 15; Mismatches 54; Indels 11; Gaps 4;
QY 35 APPSAAAPDSGPEAGLHSGMLEDLGSPNGVPRSTAP--GGIPNPEKKTNCETQCPNPQS 92
Db 791 APENAA-----GRQGETLEGDMVRGLPSPDAQNATAPVREGLPADPIARN--VRMPQPES 844
QY 93 LSSGPLETKQKNGIQTTEAKRDA---KMPAKEVTINVTDSIQMDRSRRRI 139
Db 845 LPRTVRDSLPELAQAEVRRQAGNRDIPQETIAPESSETTVSTDREATV 894
RESULT 14
D85644
Hypothetical protein Z1495 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: D85644
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85644
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2806 <STO>
A:Cross-references: UNIPROT:Q8X470; GB:AE005174; NID:g12514354; PIDN:AGS5616.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1495
Query Match 10.5%; Score 81.5; DB 2; Length 2806;
Best Local Similarity 27.3%; Pred. No. 1.4e+02;
Matches 30; Conservative 15; Mismatches 54; Indels 11; Gaps 4;
QY 35 APPSAAAPDSGPEAGLHSGMLEDLGSPNGVPRSTAP--GGIPNPEKKTNCETQCPNPQS 92
```

```
Db 804 APENAA-----GRQGETLEGDMVRGLPSPDAQNATAPVREGLPADPIARN--VRMPQPES 857
QY 93 LSSGPLETKQKNGIQTTEAKRDA---KMPAKEVTINVTDSIQMDRSRRRI 139
Db 858 LPRTVRDSLPELAQAEVRRQAGNRDIPQETIAPESSETTVSTDREATV 907
RESULT 15
A85359
translation initiation factor-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: A85359
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: A85359
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-263 <STO>
A:Cross-references: UNIPROT:Q9SUH8; GB:NC_001268; NID:g7269969; PIDN:CAB79786.1; GSPDB:G
C:Genetics:
A:Gene: AT4g30680
A:Map position: 4
Query Match 10.3%; Score 80.5; DB 2; Length 263;
Best Local Similarity 28.1%; Pred. No. 11;
Matches 38; Conservative 13; Mismatches 57; Indels 27; Gaps 7;
QY 2 GCGGSRADAIEPRYVESWTRETESTWLTYTSDAPPSAAAPDSGPEAGLHSGMLEDG-- 59
Db 15 GRGGSRRFA--PRP-----TLSSSDLTNGGDAPSPAVKSGS-----GLLNDRPS 57
QY 60 --LFSNGVPRSTAPGGIPNPEKKTNCETQCPNPQSLSSGPLTKQKNGIQTTEAKRDAKRM 117
Db 58 ALVQNG--SQQPKPVPSPTRQT-VEKPKPQPQEQEVAPPT--TTSLNTVELSRKTNLSL 111
QY 118 PAKEVTINVTDSIQQ 132
Db 112 LEEYFNVRLLDEALQ 126
Search completed: November 17, 2004, 15:28:53
Job time : 21.6667 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 15:02:57 ; Search time 24.3333 Seconds  
(without alignments)  
395.183 Million cell updates/sec

Title: US-10-705-716A-8

Perfect score: 767

Sequence: 1 MCGGSRADAEIPRYESWT.....VTENIRQMDRSKRVTKNCIN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- Issued Patents\_AA.\*
- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pap.\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pap.\*
  - 5: /cgn2\_6/ptodata/1/iaa/PTCUTUS\_COMB.pap.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87.5	11.4	778	6	5198347-4 Patent No. 5198347
2	80.5	10.5	2516	3	US-08-374-077C-2 Sequence 2, Appli
3	80.5	10.5	2516	3	US-08-895-590-2 Sequence 2, Appli
4	80.5	10.5	2516	4	US-09-539-879A-2 Sequence 2, Appli
5	79	10.3	244	4	US-09-252-991A-18465 Sequence 1, Appli
6	78	10.2	724	3	US-09-121-964-1 Sequence 1, Appli
7	77	10.0	715	4	US-09-252-991A-32740 Sequence 1, Appli
8	75.5	9.8	286	4	US-09-071-035-176 Sequence 176, App
9	75.5	9.8	305	4	US-09-071-035-174 Sequence 174, App
10	75.5	9.8	700	3	US-09-408-647A-2 Sequence 2, Appli
11	74	9.6	726	4	US-09-328-352-8115 Sequence 8115, Ap
12	74	9.6	1113	4	US-09-618-435-9 Sequence 9, Appli
13	73.5	9.6	346	4	US-09-538-092-723 Sequence 723, App
14	73	9.5	316	4	US-09-248-796A-16753 Sequence 16753, A
15	73	9.5	512	4	US-09-496-320-11 Sequence 11, Appl
16	72.5	9.5	1050	4	US-09-538-092-131 Sequence 131, App
17	72	9.4	386	4	US-09-328-352-7679 Sequence 7679, Ap
18	72	9.4	853	3	US-09-254-325-2 Sequence 2, Appli
19	72	9.4	866	3	US-09-073-415-6 Sequence 6, Appli
20	72	9.4	866	3	US-08-750-458A-2 Sequence 2, Appli
21	71.5	9.3	240	4	US-09-252-991A-21991 Sequence 21991, A
22	71.5	9.3	258	4	US-09-270-767-43579 Sequence 43579, A
23	71.5	9.3	384	4	US-09-270-767-42538 Sequence 42538, A
24	71.5	9.3	3623	4	US-09-341-461-2 Sequence 2, Appli
25	71	9.3	224	4	US-09-270-767-59848 Sequence 59848, A
26	71	9.3	433	4	US-09-270-767-44417 Sequence 44417, A
27	71	9.3	754	4	US-09-252-991A-18938 Sequence 18938, A

28	70.5	9.2	172	4	US-09-252-991A-17765 Sequence 17765, A
29	70.5	9.2	226	4	US-09-252-991A-30546 Sequence 30546, A
30	70	9.1	217	4	US-09-252-991A-26962 Sequence 26962, A
31	70	9.1	266	4	US-09-252-991A-27673 Sequence 27673, A
32	70	9.1	360	3	US-09-509-902A-9 Sequence 9, Appli
33	70	9.1	546	4	US-09-235-153-3 Sequence 3, Appli
34	70	9.1	662	4	US-09-543-681A-7467 Sequence 7467, Ap
35	69.5	9.1	139	4	US-09-252-991A-19245 Sequence 19245, A
36	69.5	9.1	219	4	US-09-270-767-57118 Sequence 57118, A
37	69.5	9.1	323	4	US-09-270-767-41874 Sequence 41874, A
38	69	9.0	245	4	US-09-252-991A-19505 Sequence 19505, A
39	69	9.0	551	2	US-09-033-537A-1 Sequence 1, Appli
40	69	9.0	709	1	US-08-444-005-17 Sequence 17, Appl
41	69	9.0	803	1	US-08-158-232-10 Sequence 10, Appl
42	69	9.0	803	1	US-08-304-626-10 Sequence 10, Appl
43	69	9.0	803	1	US-08-316-301A-12 Sequence 12, Appl
44	69	9.0	803	2	US-08-611-928-10 Sequence 10, Appl
45	69	9.0	803	3	US-09-173-891-10 Sequence 10, Appl

## ALIGNMENTS

RESULT 1  
5198347-4  
; Patent No. 5198347  
; APPLICANT: MILLER, LOUIS H.; ADAMS, JOHN H.; KASLOW, DAVID C.; FANG, XIANGDOUG  
; TITLE OF INVENTION: DNA ENCODING PLASMODIUM VIVAX AND PLASMODIUM KNOWLESII DUFFY RECEPTOR  
; NUMBER OF SEQUENCES: 27  
; CURRENT APPLICATION NUMBER: US/08/554,837  
; FILING DATE: 20-JUL-1990  
; SEQ ID NO: 4:  
; LENGTH: 778  
5198347-4

Query Match 11.4%; Score 87.5; DB 6; Length 778;  
Best Local Similarity 31.8%; Pred. No. 0.4;  
Matches 28; Conservative 13; Mismatches 44; Indels 3; Gaps 3;  
QY 31 TSDALPSAAATDSGPEAGGLHAGVLEDGLSNGV-LRPAAPGGIANPEKXNC-GTQCP 88  
DB 352 TVSSDVSFVGKDSGSPSTSASHALAGENGVEHNGTDTPEKDEKADFKQIEVKGQDT 411  
QY 89 NSQNLSS-GPITQKQNGLMWATEAKRDAK 115  
DB 412 DDRSQSLGPHTDERTATIGTETHMEKDT 439

RESULT 2  
US-08-374-077C-2  
; Sequence 2, Application US/08374077C  
; Patent No. 6027912  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Linda M.  
; APPLICANT: Ren, Dejian  
; APPLICANT: Zheng, Wei  
; APPLICANT: Dubald, Manuel Marcel Paul  
; TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha  
; TITLE OF INVENTION: Calcium Channel Subunit  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: BURNS, DOANE, SWECKER & MATHIS, LLP  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,077C
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-374-077C-2
;
; Query Match 10.5%; Score 80.5; DB 3; Length 2516;
; Best Local Similarity 22.4%; Pred. No. 14;
; Matches 39; Conservative 23; Mismatches 59; Indels 53; Gaps 8;
;
; QY 3 CGGSRADAEPRY--YESWTRETESTWLTYSALPSAAATDSGPAGGLHAGVLEDGL 60
; DB 207 CGGGISAPPPRLTPEEAQIQPQ-----NSVTSAGSTNSFSFGG-----GR 249
;
; QY 61 SSGVLRPAAPGGIANPEKQKNC-----GTQC-----PNS 90
; DB 250 DNSSY--SAVGDDSSSSNSCNDITGDNSTLHGLGVGVCFTADCDNSEDGDPNN 307
;
; QY 91 QNLSSGPL-TQKQGLWATEAKRDAKMSAREVAINTENIRQMDRSKRVTKNC 143
; DB 308 QDLSSQTLRTAAIVAATAAAKEAQEQSLADCE-SFSDRRQDADEDVRIQDC 360
;
; RESULT 3
; US-08-895-590-2
; Sequence 2, Application US/08895590
; Patent No. 6207410
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,590
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,888
; FILING DATE: 19-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-263
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-895-590-2
;
; Query Match 10.5%; Score 80.5; DB 3; Length 2516;
; Best Local Similarity 22.4%; Pred. No. 14;
; Matches 39; Conservative 23; Mismatches 59; Indels 53; Gaps 8;
;
; QY 3 CGGSRADAEPRY--YESWTRETESTWLTYSALPSAAATDSGPAGGLHAGVLEDGL 60
; DB 207 CGGGISAPPPRLTPEEAQIQPQ-----NSVTSAGSTNSFSFGG-----GR 249
;
; QY 61 SSGVLRPAAPGGIANPEKQKNC-----GTQC-----PNS 90
; DB 250 DNSSY--SAVGDDSSSSNSCNDITGDNSTLHGLGVGVCFTADCDNSEDGDPNN 307
;
; QY 91 QNLSSGPL-TQKQGLWATEAKRDAKMSAREVAINTENIRQMDRSKRVTKNC 143
; DB 308 QDLSSQTLRTAAIVAATAAAKEAQEQSLADCE-SFSDRRQDADEDVRIQDC 360
;
; RESULT 4
; US-09-539-879A-2
; Sequence 2, Application US/09539879A
; Patent No. 6436627
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Invertebrate Algal
; Calcium Channel Subunit
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/539,879A
; FILING DATE: 31-Mar-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/111,865
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/374,077
; FILING DATE: 19-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-895-590-2

```

;  
US-09-539-879A-2

Query Match	10.5%;	Score	80.5;	DB	4;	Length	2516;
Best Local Similarity	22.4%;	Pred.	No.14;				
Matches	39;	Conservative	23;	Mismatches	59;	Indels	53; Gaps
							8;
QY	3	CGGSRADAIEPRY--YESWTRFTESTWLVTYDSDALPSAAATDSGPAGGLHAGVLEDGL	60				
			:	:	:	:	:
			:	:	:	:	:
			:	:	:	:	:
Db	207	CGGGTSAPPRLTPEAWQLQP-----NSVTSAGSTNSFSFG-----GR	249				
			:	:	:	:	:
			:	:	:	:	:
QY	61	SSNGVLRPAPGGIANPERKMC-----GTQC-----PNS	90				
			:	:	:	:	:
			:	:	:	:	:
Db	250	DDNSSY--SAVGGDSSSSNSNCNDITGNSTLHGLGVDCSFADCDNSEDDEDDGDPPNN	307				
			:	:	:	:	:
			:	:	:	:	:
QY	91	QNLSGGPL-TQKONGLWATEAKRDAKMWSAREVAINTENIRQMDRSKRVTYKNC	143				
			:	:	:	:	:
			:	:	:	:	:
Db	308	QDLSSOTLRTPAATAIVAAAAAKEQAQBSLADCE-SFSDRRQDADEVRIODC	360				
			:	:	:	:	:
			:	:	:	:	:

```

RESULT 5
US-09-252-991A-18465
; Sequence 18465, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.1136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18465
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18465

```

Query Match	10.3%; Score 79; DB 4; Length 244;
Best Local Similarity	24.0%; Pred.No. 0.72;
Matches	42; Conservative 23; Mismatches 74; Indels 36; Gaps 6;

  

QY	2	GCGGSRADALEPRYESWTRETESTWLT-----YTDSALPSAAATDGSPEAGGLHGVLT	56
Db	74	GCAQARPARPGRARNRSSETFAARSPPSPGTHSSRLPARAGGTSSSIQTASRLA	133
		:           :           :           :	
QY	57	EDGLSNGVLRLPAAPGGIANPEKKMNC-----GTQCPNSQN--	92
		:           :           :           :	
Db	134	GAFRLNLG--RPSASSGIASPAPHCARPQAHHQRPRPANSSQSRRATARDNSNPA	191
		:           :           :           :	
QY	93	--LSSGPLTKOKNLWATEAKRDKNRSAREVAINTENIQMDRSKRVTKNCIN	145
		:           :           :           :	
Db	192	CRLASGSTKTSTSVASSGSKPWRRKRROTSLRT--LRQART--GTSPKAIN	243
		:           :           :           :	

```

RESULT 6
US-09-121-964-1
; Sequence 1, Application US/09121964
; Patent No. 6124447
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NOVEL ENZYME CATALYZING DEPHOSPHORYLATION
; FILE REFERENCE: 32290-144753
; CURRENT APPLICATION NUMBER: US/09/121.964
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 724
;

```

```

; TYPE: PR1
; ORGANISM: Sarcophaga peregrina
US-09-121-964-1

Query Match 10.2%; Score 78; DB 3; Length 724;
Best Local Similarity 27.7%; Pred.No. 4.6;
Matches 31; Conservative 8; Mismatches 39; Indels 34; Gaps 4;

Qy 25 STWLTVYDSDALPSAATDGP--EAGLLHAGVLEDGL-----SSNGV 65
||| : ||| : ||| : ||| : ||| : ||| : |||
Db 2 STTIEVYQDLHLLEIFKPPDAGSCEKGSLEEDITNRCEFLFKDYTLIELDNSGV 61
||| : ||| : ||| : ||| : ||| : ||| : |||

Qy 66 LRPAAGGGTANP-----KMGCGTCQPNSONLSSGPLTKQKNGL 105
||| : ||| : ||| : ||| : ||| : ||| : |||
Db 62 LSPRYGRIPIPEYEHGHTSTITPNTNLGTQATNHASPSQG---KNGNGL 110
||| : ||| : ||| : ||| : ||| : ||| : |||

RESULT 7
US-09-252-991A-32740
; Sequence 32740; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32740
; LENGTH: 715
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32740

```

Query Match	10.0%	Score 77;	DB 4;	Length 715;	
Best Local Similarity	17.9%;	Pred. No. 5.8;			
Matches	30;	Conservative 25;	Mismatches 43;	Indels 70;	Gaps 5;
Qy	39	AAATDGSPEAGLHAGVLEDGL	-----SSNGVL-----	66	
Db	417	AAARKSAVQLQGHGADAEAGLHNAVQQAQGLAGGELLWLHGQPHVPRPSEFFA	476		
Qy	67	-----RPAAPGGIANPEKMKNCGTQCPNSQLSSGPLTKOKGLLWATEAKRD	113		
Db	477	PQRSCSHKHKEGEPSCPCQLQPLASRINAST	-----ASAPSAASANGLTSSSSKRS	528	
Qy	114	A-----KRMAREVAIVNTEN-----	IRMDRSKRVTKNCI	144	
Db	529	ACQSAQSATARAATAATRASMHITAGRPKPLNSGALRRLSRAASRAAASVV	576		

```

RESULT 8
US-09-071-035-176
; Sequence 176, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

```





```

QY 52 HAGVLEGLSSNGVLRPAAPGGIANPEKKMNC--GTQCPNSQNLSSGPI---TQKQGLW 101
Db 914 QLNLLQEQMPAPSEADAVRRG-AGPDAKHCVTGPSGRSRHRCTSGELATATAHQESAA 972
QY 107 ATEAKRDARMSAREVAINTEN-IRQWDRSK 137
Db 973 ASGSSASSIYFSTDYASEVSENRRPQDRQR 1004

RESULT 13
US-09-538-092-723
; Sequence 723, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 723
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YOR007C
US-09-538-092-723

Query Match 9.6%; Score 73.5; DB 4; Length 346;
Best Local Similarity 29.4%; Pred. No. 5.2;
Matches 37; Conservative 11; Mismatches 51; Indels 27; Gaps 6

QY 5 GSPADAIEPRYVESWTRETE-STWLTYT-----DSDALPSAATDSG-PPAGGHAGVL 56
Db 201 GDNATEAMKRDYESAKKKVEQSLNLEKTVPEQSRDADVDASQGSAGGLPDLGSLGGGL 260
QY 57 EDGLSSNGVLRPAA-----PGGIANPEKKMNC-----CGTQCPNSQNLSSGPI 98
Db 261 -GGLMNPQLMQAAQKQMSNPGAMQNIQKMQDPFSIQMAEGFASGGTPTNLSIDMNP 319
QY 99 TQKQNG 104
Db 320 LRNMAG 325

RESULT 14
US-09-248-796A-16753
; Sequence 16753, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDATE
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16753
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Candida albicans

```

US-09-248-796A-16753

Query Match 9.5%; Score 73; DB 4; Length 316;  
Best Local Similarity 34.7%; Pred. No. 5.2; Mismatches 9; Indels 8; Gaps 4;  
Matches 26; Conservative 9; Mismatches 9; Indels 8; Gaps 4;  
Qy 57 EDGLSNGVLRPAAPGGIANPEKMKMC-GTCCPNQNL-SSGPILOKQGLWATEAKRD- 113  
Db 74 EDNFTQSTDTTAAAVAKSNP--KTNAEPKIPNEKLLKTESPLSQXQNGATTTKESDV 131  
Qy 114 ---AKRMSAREVAIN 125  
Db 132 LLETSTSTSTVSNN 146

RESULT 15

US-09-496-320-11  
; Sequence 11, Application US/09496320  
; Patent No. 6593080  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Alvin  
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of Calicivirus  
; FILE OF INVENTION: Infection in Humans  
; FILE REFERENCE: 54184  
; CURRENT APPLICATION NUMBER: US/09/496,320  
; CURRENT FILING DATE: 2000-02-01  
; EARLIER APPLICATION NUMBER: 60/118,209  
; EARLIER FILING DATE: 1999-02-01  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 512  
; TYPE: PRT  
; ORGANISM: Caliciviridae  
US-09-496-320-11

Query Match 9.5%; Score 73; DB 4; Length 512;  
Best Local Similarity 24.4%; Pred. No. 10;  
Matches 40; Conservative 17; Mismatches 57; Indels 50; Gaps 8;  
Qy 2 GCGSRADAIEPRYY-----ESWTRETESTWLTYSDD---ALPSAAATDSGPBAGG 50  
Db 152 GCGSKMSHFPHRDHGKQPAHRGSSWTRH-QHTQPSVENADYPHGLRAQSPQNGRHAFG 210  
Qy 51 LHAGVLEDGLSSNGVLRPA--APGGIANPEK-KMNCGT-----QCPNSQNLSSGPILO 100  
Db 211 LHAESHRLRPLHRRPEKPCGRCPLAGPPQFQKMEPCGLSSTTRVCPRTS-----TK 266  
Qy 101 KQGLWATEAKRDAKMSAREVAINVTENIRQMDRSKRVTKNCI 144  
Db 267 QSRRAA-----QVDLGLRCWEHCL 286

Search completed: November 17, 2004, 15:30:04  
Job time : 25.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 14:58:39 ; Search time 98 Seconds  
(without alignments)  
530.773 Million cell updates/sec

Title: US-10-705-716a-8

Perfect score: 767

Sequence: 1 MCGGSRADAIERYESWT.....VTENIQMDRSKRVTKNCIN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	767	100.0	145	8	ADO48479	Ado48479 Mouse PTH
2	746	97.3	145	8	ADO48473	Ado48473 Rat PTH r
3	649	84.6	145	4	AAB95018	Aab95018 Human pro
4	649	84.6	145	5	AAO19498	Aao19498 HSL prote
5	649	84.6	145	6	ABR58646	Abf58646 Human can
6	649	84.6	145	7	ADC31800	Adc31800 Human nov
7	649	84.6	145	7	ADM46959	Adm46959 Brain and
8	649	84.6	145	8	ADO48475	Ado48475 Human PTH
9	621.5	81.0	180	7	ADM46961	Adm46961 Brain and
10	472.5	61.6	149	7	ADM46963	Adm46963 Brain and
11	329.5	43.0	92	5	AQO81902	Aqo81902 Human dio
12	294	38.3	54	8	ADO48481	Ado48481 Mouse PTH
13	288	37.5	73	7	ADM46962	Adm46962 Brain and
14	278	36.2	54	7	ADM46960	Adm46960 Brain and
15	278	36.2	54	8	ADO48477	Ado48477 Human PTH
16	272	35.5	80	7	ADM46964	Adm46964 Brain and
17	140	18.3	25	7	ADM46979	Adm46979 Brain and
18	99	12.9	18	7	ADM46977	Adm46977 Brain and
19	89	11.6	16	8	ADO48482	Ado48482 PTH respo
20	87.5	11.4	778	2	AA13456	Aa13456 Duffy rec
21	84.5	11.0	684	4	ABR69330	Abf69330 Drosophil
22	84	11.0	1433	5	ABP35624	Abp35624 Fungal ZB
23	82	10.7	574	5	ABG95028	Abg95028 Human tra
24	81.5	10.6	1100	4	AAG84930	Aag84930 Shrimp wh
25	80.5	10.5	600	7	ADJ70139	Adj70139 Human hea

## ALIGNMENTS

### RESULT 1

ADO48479  
ID ADO48479 standard; protein; 145 AA.

XX AC ADO48479;

XX DT 12-AUG-2004 (first entry)

XX DE Mouse PTH responsive gene protein.

XX KW PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;

XX KW transgenic animal; osteopathic; gene therapy; osteoporosis; mouse;

XX KW murine.

XX OS Mus sp.

XX FN WO2004044152-A2.

XX PD 27-MAY-2004.

XX PF 10-NOV-2003; 2003WO-US035655.

XX PR 12-NOV-2002; 2002US-0425532P.

XX PA (AMHP ) WYETH.

XX PI Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;

XX DR WPI; 2004-420299/39.

XX DR N-PSDB; ADO48478.

PT New nucleic acid fragment encoding a PAIGB polypeptide, useful in

PT preparing a composition for diagnosing, treating or preventing bone

PT related disorders, e.g., osteoporosis.

XX PS Claim 9; SEQ ID NO 8; 169pp; English.

XX CC The invention relates to a novel PTH responsive gene (PAIGB) fragment

XX CC encoding a polypeptide. The invention further comprises: a chimeric

XX CC construct comprising the isolated nucleic acid fragment operatively

XX CC linked to suitable regulatory sequences; a host cell transformed with the

XX CC chimeric construct; a vector comprising the nucleic acid fragment;

XX CC obtaining a nucleic acid fragment encoding the polypeptide; a method for

XX CC obtaining a polypeptide; detecting the presence of the nucleic acid

XX CC fragment; an antibody that specifically binds to one or more epitopes of

XX CC a PAIGB polypeptide; a composition for regulating bone-forming activity

XX CC in a mammal comprising the nucleic acid fragment, polypeptide or antibody

XX CC ; an agent that alters the expression of PAIGB gene or polypeptide;

XX CC

CC determining whether an agent alters the expression of PAIGB mRNA;  
 CC screening agents for effectiveness in altering expression of the nucleic  
 CC acid fragment; screening for agents useful for treating bone related  
 CC disorders; evaluating the efficacy of a treatment of a bone related  
 CC disorder in a subject; identifying polypeptides capable of binding to  
 CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
 CC related agent; a transgenic animal comprising the DNA; an animal model  
 CC for the study of bone density modulation comprising a first group of  
 CC animals composed of the transgenic animal and a second group of control  
 CC animals; studying bone mass determinants; studying the modulation of bone  
 CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
 CC for treating bone related disorders; identifying whether an agent which  
 CC has bone forming activity; and a stably transfected cell line comprising  
 CC two constructs, the first construct comprising a ligand binding domain  
 CC linked to a DNA binding domain which is linked to an activation domain  
 CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

XX Sequence 145 AA;

Query Match 100.0%; Score 767; DB 8; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-78;  
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCGGSRADAEIPRYESWTRETESTWLTVDSDALPSAAATSGPPAGGLHAGVLEDGL 60  
 Db 1 MCGGSRADAEIPRYESWTRETESTWLTVDSDALPSAAATSGPPAGGLHAGVLEDGL 60

Qy 61 SSGVLRPAAPGGIANPEKKNCGTCQPNQNLSSGGLTKQKGLWATEAKRDKRMSAR 120  
 Db 61 SSGVLRPAAPGGIANPEKKNCGTCQPNQNLSSGGLTKQKGLWATEAKRDKRMSAR 120

Qy 121 EVAINVTENIRQMDRSKRVTNCIN 145  
 Db 121 EVAINVTENIRQMDRSKRVTNCIN 145

RESULT 2  
 ADO48473  
 ID ADO48473 standard; protein; 145 AA.

XX ADO48473;  
 XX 12-AUG-2004 (first entry)  
 XX Rat PTH responsive gene protein.  
 XX PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
 XX transgenic animal; osteopathic; gene therapy; osteoporosis; rat.  
 XX Rattus sp.  
 XX W0200404152-A2.  
 XX 27-MAY-2004.  
 XX 10-NOV-2003; 2003WO-US035655.  
 XX 12-NOV-2002; 2002US-0425532P.  
 XX (AMHP ) WYETH.  
 XX Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;  
 XX WPI; 2004-420299/39.  
 XX N-ESDB; ADO48472.

XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
 PT preparing a composition for diagnosing, treating or preventing bone  
 PT related disorders, e.g., osteoporosis.

XX Claim 9; SEQ ID NO 2; 169pp; English.

XX The invention relates to a novel PTH responsive gene (PAIGB) fragment  
 CC encoding a polypeptide. The invention further comprises: a chimeric  
 CC construct comprising the isolated nucleic acid fragment operatively  
 CC linked to suitable regulatory sequences; a host cell transformed with the  
 CC chimeric construct; a vector comprising the nucleic acid fragment;  
 CC obtaining a nucleic acid fragment encoding the polypeptide; a method for  
 CC fragmenting a polypeptide; detecting the presence of the nucleic acid  
 CC fragment; an antibody that specifically binds to one or more epitopes of  
 CC a PAIGB polypeptide; a composition for regulating bone-forming activity  
 CC in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
 CC ; an agent that alters the expression of PAIGB gene or polypeptide;  
 CC determining whether an agent alters the expression of PAIGB mRNA;  
 CC screening agents for effectiveness in altering expression of the nucleic  
 CC acid fragment; screening for agents useful for treating bone related  
 CC disorders; evaluating the efficacy of a treatment of a bone related  
 CC disorder in a subject; identifying polypeptides capable of binding to  
 CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
 CC related agent; a transgenic animal comprising the DNA; an animal model  
 CC for the study of bone density modulation comprising a first group of  
 CC animals composed of the transgenic animal and a second group of control  
 CC animals; studying bone mass determinants; studying the modulation of bone  
 CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
 CC for treating bone related disorders; identifying whether an agent which  
 CC has bone forming activity; and a stably transfected cell line comprising  
 CC two constructs, the first construct comprising a ligand binding domain  
 CC linked to a DNA binding domain which is linked to an activation domain  
 CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

XX Sequence 145 AA;

Query Match 97.3%; Score 746; DB 8; Length 145;  
 Best Local Similarity 97.2%; Pred. No. 6.3e-76;  
 Matches 141; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MCGGSRADAEIPRYESWTRETESTWLTVDSDALPSAAATSGPPAGGLHAGVLEDGL 60  
 Db 1 MCGGSRADAEIPRYESWTRETESTWLTVDSDALPSAAATSGPPAGGLHAGVLEDGP 60

Qy 61 SSGVLRPAAPGGIANPEKKNCGTCQPNQNLSSGGLTKQKGLWATEAKRDKRMSAR 120  
 Db 61 SSGVLRPAAPGGIANPEKKNCGTCQPNQNLSSGGLTKQKGLWATEAKRDKRMSAR 120

Qy 121 EVAINVTENIRQMDRSKRVTNCIN 145  
 Db 121 EVAINVTENIRQMDRSKRVTNCIN 145

RESULT 3  
 AAB95018  
 ID AAB95018 standard; protein; 145 AA.  
 XX AAB95018;  
 XX 26-JUN-2001 (first entry)  
 XX Human protein sequence SEQ ID NO:16726.  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

```

XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX PT length cDNAs defined in the specification, and for the detection and/or
XX PT diagnosis of the abnormality of the proteins encoded by the full-length
XX PT cDNAs.
XX PS Claim 8; SEQ ID NO 16726; 2537pp + Sequence Listing; English.
XX CC The present invention describes primer sets for synthesizing 5602 full-
XX CC length cDNAs defined in the specification. Where a primer set comprises:
XX CC (a) an oligo-dT primer and an oligonucleotide complementary to the
XX CC complementary strand of a polynucleotide which comprises one of the 5602
XX CC nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in the
XX CC specification. The primer sets can be used in antisense therapy and in
XX CC gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
XX CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX CC present invention
XX SQ Sequence 145 AA;

Query Match 84.6%; Score 649; DB 4; Length 145;
Best Local Similarity 83.4%; Pred. No. 6e-65;
Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 MCGCGSRADAIEPRYESWTRETSTWLTYSALPSAAATDSGPEAGGLHAGVLEDGL 60
Db 1 MCGCGSRADAIEPRYESWTRETSTWLTYSALPSAAATDSGPEAGGLHAGVLEDGL 60

QY 61 SSGVLRPAAPGGTANPEKKNCCGCPNSQNLSSGGLTKQKGLWATEAKRDKRMGAR 120
Db 61 PSNGVPRSTAPGGTANPEKKNCCGCPNSQNLSSGGLTKQKGLWATEAKRDKRMGAR 120

QY 121 EVAINVTENIRQMDRSKRVTNKCIN 145
Db 121 EVTINVTDISIQQMDRSRRTIKNCVN 145

RESULT 4
AAO19498

```

```

ID AAO19498 standard; protein; 145 AA.
XX AC AAO19498;
XX DT 20-DEC-2002 (first entry)
XX DE HSI protein variant.
XX KW HSI; variant; cancer; tumour; unigene cluster; cytostatic; metastasis;
KW EST; expressed sequence tag; colon cancer; stomach cancer; breast cancer;
KW HSI69395; HSI27144; HS2; HSI32793; HS3.
XX OS Unidentified.
XX PN DE10103694-A1.
XX PD 01-AUG-2002.
XX PF 26-JAN-2001; 2001DE-01003694.
XX PR 26-JAN-2001; 2001DE-01003694.
XX PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.
XX PI Brett D, Kemmer W;
XX DR WPI; 2002-644836/70.
XX DR N-PSDB; AAL50100.
XX PT Diagnosis and therapy of tumors, by determining expression rates of
XX PT specific expressed sequence tags of the unigene cluster, and subsequently
XX PT blocking their expression.
XX PS Claim 10; Page 5; 10pp; German.
XX CC The present invention relates to the use of expressed sequence tags
XX CC (ESTs), or variants, of the unigene cluster HSI69395 (HS1), HSI27144
XX CC (HS2) and/or HSI32793 (HS3) for diagnosis and therapy of tumours, in
XX CC which their expression rates in tumour cells and/or lymph nodes are
XX CC determined. The EST sequences are useful as prognostic markers of
XX CC survival of cancer patients (high levels of EST-related mRNA are
XX CC associated with a poor prognosis, specifically correlated with
XX CC development of metastases); and for diagnosis and/or therapy of solid
XX CC tumours, particularly of colon, stomach and breast. The present sequence
XX CC is a variant of the HSI protein shown in the exemplification of the
XX CC invention
XX SQ Sequence 145 AA;

Query Match 84.6%; Score 649; DB 5; Length 145;
Best Local Similarity 83.4%; Pred. No. 6e-65;
Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 MCGCGSRADAIEPRYESWTRETSTWLTYSALPSAAATDSGPEAGGLHAGVLEDGL 60
Db 1 MCGCGSRADAIEPRYESWTRETSTWLTYSALPSAAATDSGPEAGGLHAGVLEDGL 60

QY 61 SSGVLRPAAPGGTANPEKKNCCGCPNSQNLSSGGLTKQKGLWATEAKRDKRMGAR 120
Db 61 PSNGVPRSTAPGGTANPEKKNCCGCPNSQNLSSGGLTKQKGLWATEAKRDKRMGAR 120

QY 121 EVAINVTENIRQMDRSKRVTNKCIN 145
Db 121 EVTINVTDISIQQMDRSRRTIKNCVN 145

RESULT 5
ABR58646
ID ABR58646 standard; protein; 145 AA.
XX AC ABR58646;
XX DT 09-JUL-2003 (first entry)

```

XX DE Human cancer related protein SEQ ID NO:303.

XX KW Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;

XX KW heart disease; atherosclerosis; endometriosis.

XX OS Homo sapiens.

XX PN WO2003025138-A2.

XX PD 27-MAR-2003.

XX PF 17-SEP-2002; 2002WO-US029560.

XX PR 17-SEP-2001; 2001US-0323469P.

XX PR 20-SEP-2001; 2001US-0323887P.

XX PR 13-NOV-2001; 2001US-0350666P.

XX PR 08-FEB-2002; 2002US-0355145P.

XX PR 08-FEB-2002; 2002US-0355257P.

XX PR 12-APR-2002; 2002US-0372246P.

XX PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX PF Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;

XX PI Zlotnik A;

XX DR WPI; 2003-354600/33.

XX DR N-PSDB; ACC72796.

XX PT New genes that are up-regulated or down-regulated in cancers, useful as

XX PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as

XX PT therapeutic targets for screening drugs for treating these diseases.

XX PS Claim 12; Page 753; 767pp; English.

XX CC The present invention describes an isolated nucleic acid molecule, which

CC comprises the sequence of any of the genes that are up-regulated or down-

CC regulated in specific cancers (e.g. about 1031 genes up-regulated in

CC acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer

CC related gene nucleotide sequences which encode the proteins given in

CC ARS5821 to ARS58799. Also described: (1) determining the presence or

CC absence of a pathological cell in a patient; (2) an expression vector

CC comprising a nucleic acid molecule described above; (3) a host cell

CC comprising the vector; (4) an isolated polypeptide, which is encoded by

CC the nucleic acid; (5) an antibody that specifically binds the polypeptide

CC of (4); (6) specifically targeting a compound to a pathological cell in a

CC patient by administering to the patient the antibody above; and (7) a

CC drug screening assay. The nucleic acid is useful as diagnostic markers or

CC therapeutic targets. In particular, the nucleic acid is useful for

CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,

CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,

CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,

CC atherosclerosis and endometriosis. The nucleic acid is also useful in

CC drug screening, particularly for identifying agents for treating these

CC pathologies

XX SQ Sequence 145 AA;

Query Match 84.6%; Score 649; DB 6; Length 145;

Best Local Similarity 83.4%; Pred. No. 6e-65;

Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 MCGGSRADAEPRYIESWTRFETSTWLTYSALPSAAATDSGPPAGGLHAGVLDGL 60

DB 1 MCGGSRADAEPRYIESWTRFETSTWLTYSALPSAAATDSGPPAGGLHAGVLDGL 60

QY 61 SSGVLRPAAPGGIANPEKKMNCCTQCPNSQNLSSGGLTQKQNGLWATEAKRDKRMSAR 120

DB 61 PSNGVPRSTAPGGIPNPEKKTKNCETQCPNQSLSGGLTQKQNGLWATEAKRDKRMPAK 120

QY 121 EVATNVNTFQMDRSKEVTKNCIN 145

DB 121 EVTINVTDSTQMDRSRRITKNCVN 145

RESULT 6

ADC31800

ID ADC31800 standard; protein; 145 AA.

XX AC ADC31800;

XX AC AC

XX 18-DEC-2003 (first entry)

XX DE Human novel polypeptide sequence, SEQ ID NO:1882.

XX KW Human; diagnostic; drug screening; forensics; gene mapping;

XX KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;

XX KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;

XX KW ulcers; osteoporosis; autoimmune disease; cancer;

XX KW molecular weight marker; food supplement; antiparkinsonian; nootropic;

XX KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;

XX KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;

XX KW gene therapy; chromosome 8.

XX OS Homo sapiens.

XX PN WO2003029271-A2.

XX PD 10-APR-2003.

XX PF 24-SEP-2002; 2002WO-US030474.

XX PR 24-SEP-2001; 2001US-0324631P.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;

XX PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;

XX PI Haley-Vicente D, Drmanac RT;

XX DR WPI; 2003-371981/35.

XX DR N-PSDB; ADC30829.

XX PT New polynucleotide and polypeptide useful for diagnosing, preventing or

XX PT treating conditions such as neurodegenerative diseases, anemias, platelet

XX PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or

XX PT cancer.

XX PS Claim 20; SEQ ID NO 1882; 1185pp; English.

XX CC The invention relates to 971 novel human cDNA sequences (ADC29919-

XX CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The

XX CC invention also relates to nucleic acid sequences over 99% identical with

XX CC the novel human cDNAs. The invention additionally encompasses expression

XX CC vectors and host cells comprising a nucleic acid of the invention; the

XX CC recombinant production of a polypeptide of the invention; an antibody

XX CC against a polypeptide of the invention; a method of detecting

XX CC polynucleotides or polypeptides of the invention; and methods of

XX CC identifying a compound which binds to a polypeptide of the invention. The

XX CC invention further discloses methods of preventing, treating or

XX CC ameliorating a medical condition; kits comprising polynucleotide probes

XX CC and/or monoclonal antibodies for carrying out the methods of the

XX CC invention; methods for the identification of compounds that modulate the

XX CC expression or activity of the polynucleotide and/or polypeptide; and 767

XX CC contig sequences corresponding to the cDNA sequences of the invention

XX CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628

XX CC -ADC33394). The nucleic acids and polypeptides of the invention are

XX CC useful in diagnostics, drug screening, forensics, gene mapping, in the

XX CC identification of mutations responsible for genetic disorders or other

XX CC traits, for assessing biodiversity, and in producing many other types of

XX CC data and products dependent on DNA and amino acid sequences. They are

XX CC also used for treating diseases such as Parkinson's disease, Alzheimer's

XX CC disease and other neurodegenerative diseases, anaemia, platelet

XX CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or

XX CC cancer. The nucleic acids may also be used as hybridisation probes or

XX CC primers, and in the recombinant production of a protein. The polypeptides

CC are also useful in generating antibodies, as molecular weight markers,  
 CC and as food supplements. The present sequence represents a specifically  
 CC claimed human polypeptide sequence of the invention. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 145 AA;  
 Query Match 84.6%; Score 649; DB 7; Length 145;  
 Best Local Similarity 83.4%; Pred. No. 6e-65;  
 Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSDDALPSAAATDSGPEAGLHAGVLEDGL 60  
 DB 1 MCGGSRADAIEPRYYESWTRETESTWLTYSDDALPSAAATDSGPEAGLHAGVLEDGL 60  
 QY 61 SNGVLRPAAPGGIANPEKMKCGTCQPNQSLSSGPLTKQKGLWATEAKDKRMSAR 120  
 DB 61 PSNGVPRSTAPGGIPNPEKTKTCETQCPNPQSLSSGPLTKQKGLWATEAKDKRMPAK 120  
 QY 121 EVAINVTENIROMDRSKRVTKNCIN 145  
 DB 121 EVTINVTDSIQQMDRSRRITKNCVN 145

RESULT 7  
 ADM46959  
 ID ADM46959 standard; protein; 145 AA.  
 XX  
 AC ADM46959;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Brain and Acute Leukemia, Cytoplasmic alternate protein #1.  
 XX  
 KW acute myelogenous leukemia; gene expression; BAALC;  
 KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 KW Cytoplasmic; exon.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 41 /note= "encoded by GCS"  
 FT  
 XX  
 PN WO2003040347-A2.  
 XX  
 PD 15-MAY-2003.  
 XX  
 PF 12-NOV-2002; 2002WO-US036375.  
 XX  
 PR 09-NOV-2001; 2001US-0348210P.  
 XX  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX  
 FI Tanner SM, De La Chapell A;  
 XX  
 DR WPI; 2003-441564/41.  
 DR N-PSDB; ADM46951.  
 XX  
 PT Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 PT in a patient comprises assaying for the overexpression of one or more  
 PT BAALC transcripts in cells obtained from the patient.  
 XX  
 PS Disclosure; SEQ ID NO 17; 78pp; English.  
 XX  
 CC The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful

CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.

SQ Sequence 145 AA;

Query Match 84.6%; Score 649; DB 7; Length 145;  
 Best Local Similarity 83.4%; Pred. No. 6e-65;  
 Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSDDALPSAAATDSGPEAGLHAGVLEDGL 60  
 DB 1 MCGGSRADAIEPRYYESWTRETESTWLTYSDDALPSAAATDSGPEAGLHAGVLEDGL 60  
 QY 61 SNGVLRPAAPGGIANPEKMKCGTCQPNQSLSSGPLTKQKGLWATEAKDKRMSAR 120  
 DB 61 PSNGVPRSTAPGGIPNPEKTKTCETQCPNPQSLSSGPLTKQKGLWATEAKDKRMPAK 120  
 QY 121 EVAINVTENIROMDRSKRVTKNCIN 145  
 DB 121 EVTINVTDSIQQMDRSRRITKNCVN 145

RESULT 8  
 ADO48475  
 ID ADO48475 standard; protein; 145 AA.  
 XX  
 AC ADO48475;  
 XX  
 DT 12-AUG-2004 (first entry)  
 XX  
 DE Human PTH responsive gene protein.  
 XX  
 KW PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
 KW transgenic animal; osteopathic; gene therapy; osteoporosis; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004044152-A2.  
 XX  
 PD 27-MAY-2004.  
 XX  
 PF 10-NOV-2003; 2003WO-US035655.  
 XX  
 PR 12-NOV-2002; 2002US-0425532P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;  
 XX  
 DR WPI; 2004-420299/39.  
 DR N-PSDB; ADO48474.  
 XX  
 PT New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
 PT preparing a composition for diagnosing, treating or preventing bone  
 PT related disorders, e.g., osteoporosis.  
 XX  
 PS Claim 9; SEQ ID NO 4; 169pp; English.  
 XX  
 CC The invention relates to a novel PTH responsive gene (PAIGB) fragment  
 CC encoding a polypeptide. The invention further comprises: a chimeric  
 CC construct comprising the isolated nucleic acid fragment operatively  
 CC linked to suitable regulatory sequences; a host cell transformed with the  
 CC chimeric construct; a vector comprising the nucleic acid fragment;  
 CC obtaining a nucleic acid fragment encoding the polypeptide; a method for  
 CC obtaining a polypeptide; detecting the presence of the nucleic acid  
 CC fragment; an antibody that specifically binds to one or more epitopes of  
 CC a PAIGB polypeptide; a composition for regulating bone-forming activity  
 CC in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
 CC ; an agent that alters the expression of PAIGB gene or polypeptide;  
 CC determining whether an agent alters the expression of PAIGB mRNA;  
 CC screening agents for effectiveness in altering expression of the nucleic  
 CC acid fragment; screening for agents useful for treating bone related





PT Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
PT in a patient comprises assaying for the overexpression of one or more  
PT BAALC transcripts in cells obtained from the patient.

XX  
XX Disclosure; SEQ ID NO 21; 78pp; English.

XX  
XX The invention relates to a method of characterizing acute myelogenous  
XX leukemia (AML) in a patient by assaying for the overexpression of one or  
XX more BAALC transcripts in cells obtained from the patient, where an  
XX overexpression indicates that the patient has an aggressive form of AML.  
XX The methods, kits and probes are useful for characterizing acute or  
XX chronic myelogenous leukemia, or prostate cancer. They are also useful  
XX for detecting BAALC overexpression. This sequence corresponds to a BAALC  
XX (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
XX spliced RNA consisting of exons 1, 6 and 8.

XX Sequence 149 AA;

Query Match 61.6%; Score 472.5; DB 7; Length 149;  
Best Local Similarity 65.3%; Pred. No. 6.1e-45;  
Matches 94; Conservative 3; Mismatches 12; Indels 35; Gaps 1;

QY 1 MCGGSRADAIERPRYESWTRETESTWLTYSDDLPSAANTDGPAGGLHA----- 53

Db 1 MCGGSRADAIERPRYESWTRETESTWLTYSDDLPSAANTDGPAGGLHVSLEAEKS 60

QY 54 -----GVLEDGLSSNGVLRPAAPGGIANPEKKNCGT 85

Db 61 KIKAPTDVSDGLFSASMAFLAVFSGHMGLEDGLPSNGVPRSTAPGGIPNPKTNCET 120

QY 86 QCPNQNLSGGLTKOKNGLWATE 109

Db 121 QCPNPQSLSSGGLTKOKNGLQTTE 144

RESULT 11

ADQ81902

ID ADQ81902 standard; protein; 92 AA.

AC ADQ81902;

DT 09-SEP-2004 (first entry)

XX Human dihydrogenase 10.12.

XX Human; enzyme; dihydrogenase 10.12; malignant tumour; inflammation;

XX immunological disease; haemopathy; HIV infection.

XX Homo sapiens.

XX CN1344798-A.

XX 17-APR-2002.

XX 29-SEP-2000; 2000CN-00125495.

XX 29-SEP-2000; 2000CN-00125495.

XX (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI; 2002-509506/55.

XX N-PSDB; ADQ81901.

XX New polypeptide human dihydrogenase 10.12 and polynucleotides encoding this

XX polypeptide, useful for treating various diseases, such as malignant

XX tumors, inflammations, immunological diseases, hemopathy and HIV

XX infection.

XX Claim 1; SEQ ID NO 2; 33pp; Chinese.

XX The present invention discloses a new kind of polypeptide, human

CC dihydrogenase 10.12, polynucleotides encoding this polypeptide, a DNA  
CC recombination process to produce the polypeptide and antagonist against  
CC the polypeptide. The present invention also discloses the method of  
CC applying the polypeptide in treating various diseases, such as malignant  
CC tumors, inflammations, immunological diseases, haemopathy and HIV  
CC infection. The present sequence is the human dihydrogenase 10.12.

XX Sequence 92 AA;

Query Match 43.0%; Score 329.5; DB 5; Length 92;  
Best Local Similarity 63.4%; Pred. No. 4.7e-29;  
Matches 64; Conservative 8; Mismatches 10; Indels 19; Gaps 1;

QY 45 GPEAGGLHAGVLEDGLSSNGVLRPAAPGGIANPEKKNCGTQCPNQNLSGGLTKOKNG 104

Db 11 GPETGN-----APGIPNPKTNCETQCPNPQSLSSGGLTKOKNG 51

QY 105 LMATEAKRDAKRMSAREVAINVTENIQMDRSKRVTNCIN 145

Db 52 LQTTEAKRDAKRMPAKEVTINVTDSIQMDRSRRITKNCVN 92

RESULT 12

ADQ48481

ID ADQ48481 standard; protein; 54 AA.

AC ADQ48481;

DT 12-AUG-2004 (first entry)

XX Mouse PTH responsive gene protein exon 2 splice variant.

XX PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;

XX transgenic animal; osteopathic; gene therapy; osteoporosis; mouse;

XX murine.

XX Mus sp.

XX WO2004044152-A2.

XX 27-MAY-2004.

XX 10-NOV-2003; 2003WO-US035655.

XX 12-NOV-2002; 2002US-0425532P.

XX (AMHP ) WYETH.

XX Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;

XX WPI; 2004-420299/39.

XX N-PSDB; ADQ48480.

XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in

XX preparing a composition for diagnosing, treating or preventing bone

XX related disorders, e.g., osteoporosis.

XX Claim 9; SEQ ID NO 10; 169pp; English.

XX The invention relates to a novel PTH responsive gene (PAIGB) fragment

XX encoding a polypeptide. The invention further comprises: a chimeric

XX construct comprising the isolated nucleic acid fragment operatively

XX linked to suitable regulatory sequences; a host cell transformed with the

XX chimeric construct; a vector comprising the nucleic acid fragment;

XX obtaining a nucleic acid fragment encoding the polypeptide; a method for

XX obtaining a polypeptide; detecting the presence of the nucleic acid

XX fragment; an antibody that specifically binds to one or more epitopes of

XX a PAIGB polypeptide; a composition for regulating bone-forming activity

XX in a mammal comprising the nucleic acid fragment, polypeptide or antibody

XX ; an agent that alters the expression of PAIGB gene or polypeptide;

XX determining whether an agent alters the expression of PAIGB mRNA;

XX screening agents for effectiveness in altering expression of the nucleic

XX acid fragment; screening for agents useful for treating bone related

CC disorders; evaluating the efficacy of a treatment of a bone related  
 CC disorder in a subject; identifying polypeptides capable of binding to  
 CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
 CC related agent; a transgenic animal comprising the DNA; an animal model  
 CC for the study of bone density modulation comprising a first group of  
 CC animals composed of the transgenic animal and a second group of control  
 CC animals; studying bone mass determinants; studying the modulation of bone  
 CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
 CC for treating bone related disorders; identifying whether an agent which  
 CC has bone forming activity; and a stably transfected cell line comprising  
 CC two constructs, the first construct comprising a ligand binding domain  
 CC linked to a DNA binding domain which is linked to an activation domain  
 CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopontin activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

XX SQ Sequence 54 AA;

Query Match 38.3%; Score 294; DB 8; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-25;  
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPAGGLHAG 54  
 |||||  
 DB 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPAGGLHAG 54

RESULT 13

ADM46962  
 ID ADM46962 standard; protein; 73 AA.

AC ADM46962;

XX 03-JUN-2004 (first entry)

XX Brain and Acute Leukemia, Cytoplasmic alternate protein #4.

XX acute myelogenous leukemia; gene expression; BAALC;  
 KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 KW Cytoplasmic; exon.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 41 /note= "encoded by GCS"

XX WO2003040347-A2.

XX 15-MAY-2003.

XX 12-NOV-2002; 2002WO-US036375.

XX 09-NOV-2001; 2001US-0348210P.

XX (OHIS ) UNIV OHIO STATE RES FOUND.

XX Tanner SM, De La Chapell A;

XX WPI; 2003-441564/41.

XX N-PSDB; ADM46954.

XX Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 PT in a patient comprises assaying for the overexpression of one or more  
 PT BAALC transcripts in cells obtained from the patient.

XX Disclosure; SEQ ID NO 20; 78pp; English.

XX The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
 CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.

XX SQ Sequence 73 AA;

Query Match 37.5%; Score 288; DB 7; Length 73;  
 Best Local Similarity 91.4%; Pred. No. 1.7e-24;  
 Matches 53; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPAGGLHAGVLED 58  
 |||||  
 DB 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPAGGLHAGVLED 58

RESULT 14

ADM46960  
 ID ADM46960 standard; protein; 54 AA.

XX AC ADM46960;

XX 03-JUN-2004 (first entry)

XX Brain and Acute Leukemia, Cytoplasmic alternate protein #2.

XX acute myelogenous leukemia; gene expression; BAALC;  
 KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 KW Cytoplasmic; exon.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 41 /note= "encoded by GCS"

XX WO2003040347-A2.

XX 15-MAY-2003.

XX 12-NOV-2002; 2002WO-US036375.

XX 09-NOV-2001; 2001US-0348210P.

XX (OHIS ) UNIV OHIO STATE RES FOUND.

XX Tanner SM, De La Chapell A;

XX WPI; 2003-441564/41.

XX N-PSDB; ADM46952.

XX Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 PT in a patient comprises assaying for the overexpression of one or more  
 PT BAALC transcripts in cells obtained from the patient.

XX Disclosure; SEQ ID NO 18; 78pp; English.

XX The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
 CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.

SQ Sequence 54 AA;

Query Match 36.2%; Score 278; DB 7; Length 54;  
 Best Local Similarity 94.4%; Pred. No. 1.5e-23;  
 Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCGGSRADAIEPRYESWTRETESTWLTYSALPSAAATDSGFEAGGLHAG 54  
 |||||  
 DB 1 MCGGSRADAIEPRYESWTRETESTWLTYSALPSAAATDSGFEAGGLHAG 54  
 |||||

RESULT 15  
 ADO48477  
 ID ADO48477 standard; protein; 54 AA.  
 XX AC ADO48477;  
 XX DT 12-AUG-2004 (first entry)  
 XX DE Human PTH responsive gene protein exon 2 splice variant.  
 XX KW PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
 KW transgenic animal; osteopathic; gene therapy; osteoporosis; human.  
 XX OS Homo sapiens.  
 XX PN WO200404152-A2.  
 XX PD 27-MAY-2004.  
 XX PF 10-NOV-2003; 2003WO-US035655.  
 XX PR 12-NOV-2002; 2002US-0425532P.  
 XX PA (AMHP ) WYETH.  
 XX PI Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;  
 XX DR WPI; 2004-420299/39.  
 XX DR N-PSDB; ADO48476.  
 XX PT New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
 PT preparing a composition for diagnosing, treating or preventing bone  
 PT related disorders, e.g., osteoporosis.  
 XX PS Claim 9; SEQ ID NO 6; 169pp; English.

The invention relates to a novel PTH responsive gene (PAIGB) fragment encoding a polypeptide. The invention further comprises: a chimeric construct comprising the isolated nucleic acid fragment operatively linked to suitable regulatory sequences; a host cell transformed with the chimeric construct; a vector comprising the nucleic acid fragment; obtaining a nucleic acid fragment encoding the polypeptide; a method for obtaining a polypeptide; detecting the presence of the nucleic acid fragment; an antibody that specifically binds to one or more epitopes of a PAIGB polypeptide; a composition for regulating bone-forming activity in a mammal comprising the nucleic acid fragment, polypeptide or antibody; an agent that alters the expression of PAIGB gene or polypeptide; determining whether an agent alters the expression of PAIGB mRNA; screening agents for effectiveness in altering expression of the nucleic acid fragment; screening for agents useful for treating bone related disorders; evaluating the efficacy of a treatment of a bone related disorder in a subject; identifying polypeptides capable of binding to PAIGB; monitoring the effectiveness of treatment of a subject with a bone related agent; a transgenic animal comprising the DNA; an animal model for the study of bone density modulation comprising a first group of animals composed of the transgenic animal and a second group of control animals; studying bone mass determinants; studying the modulation of bone mass; studying an effect of PAIGB on bone disorders; identifying an agent for treating bone related disorders; identifying whether an agent which has bone forming activity; and a stably transfected cell line comprising two constructs, the first construct comprising a ligand binding domain linked to a DNA binding domain which is linked to an activation domain

CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

XX SQ Sequence 54 AA;

Query Match 36.2%; Score 278; DB 8; Length 54;  
 Best Local Similarity 94.4%; Pred. No. 1.5e-23;  
 Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCGGSRADAIEPRYESWTRETESTWLTYSALPSAAATDSGFEAGGLHAG 54  
 |||||  
 DB 1 MCGGSRADAIEPRYESWTRETESTWLTYSALPSAAATDSGFEAGGLHAG 54  
 |||||

Search completed: November 17, 2004, 15:22:25  
 Job time : 99 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 15:22:38 ; Search time 81.3333 Seconds  
(without alignments)  
631.334 Million cell updates/sec

Title: US-10-705-716A-8

Perfect score: 767

Sequence: 1 MCGCGSRADAEPRYIESWT.....VTENIRQMDRSKRVTKNCIN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	649	84.6	145	14	US-10-293-239-17
2	649	84.6	145	14	US-10-177-330-30
3	621.5	81.0	140	14	US-10-293-239-19
4	472.5	61.6	149	14	US-10-293-239-21
5	288	37.5	73	14	US-10-293-239-20
6	278	36.2	54	14	US-10-293-239-18
7	272	35.5	80	14	US-10-293-239-22
8	140	18.3	25	14	US-10-293-239-37
9	99	12.9	18	14	US-10-293-239-35
10	88	11.5	394	17	US-10-425-115-236244
11	86	11.2	1597	16	US-10-437-963-180225
12	84	11.0	147	16	US-10-767-701-48905
13	84	11.0	1433	9	US-09-801-368-60

14	84	11.0	1433	15	US-10-149-310-112	Sequence 112, App
15	82.5	10.8	264	17	US-10-425-115-275422	Sequence 275422, A
16	82.5	10.8	274	15	US-10-425-114-65588	Sequence 65588, A
17	82.5	10.8	609	16	US-10-437-963-144983	Sequence 144983, A
18	82	10.7	455	15	US-10-424-599-193656	Sequence 193656, A
19	82	10.7	1479	16	US-10-437-963-180217	Sequence 180217, A
20	81	10.6	219	14	US-10-156-761-13447	Sequence 13447, A
21	80.5	10.5	600	16	US-10-408-765A-1945	Sequence 1945, App
22	80	10.4	528	16	US-10-437-963-173618	Sequence 173618, A
23	80	10.4	1108	10	US-09-949-029-108	Sequence 108, App
24	79.5	10.4	996	16	US-10-437-963-114095	Sequence 114095, A
25	79	10.3	876	9	US-09-801-368-44	Sequence 44, App1
26	78.5	10.2	1274	16	US-10-437-963-180220	Sequence 180220, A
27	78	10.2	1491	16	US-10-437-963-140197	Sequence 140197, A
28	77.5	10.1	670	14	US-10-156-761-14107	Sequence 14107, A
29	77	10.0	268	14	US-10-369-493-8240	Sequence 8240, App
30	76.5	10.0	320	15	US-10-425-114-43318	Sequence 43318, A
31	76.5	10.0	422	17	US-10-425-115-336369	Sequence 336369, A
32	76.5	10.0	426	14	US-10-369-493-3926	Sequence 3926, App
33	76	9.9	391	16	US-10-437-963-177907	Sequence 177907, A
34	75.5	9.8	276	16	US-10-437-963-158275	Sequence 158275, A
35	75.5	9.8	286	9	US-09-071-035-176	Sequence 176, App
36	75.5	9.8	286	14	US-10-206-576-176	Sequence 176, App
37	75.5	9.8	305	9	US-09-071-035-174	Sequence 174, App
38	75.5	9.8	305	14	US-10-206-576-174	Sequence 174, App
39	75.5	9.8	373	16	US-10-437-963-104701	Sequence 104701, A
40	75	9.8	167	16	US-10-767-701-46805	Sequence 46805, A
41	75	9.8	360	17	US-10-425-115-339697	Sequence 339697, A
42	75	9.8	604	16	US-10-437-963-163101	Sequence 163101, A
43	74.5	9.7	296	16	US-10-437-963-193696	Sequence 193696, A
44	74.5	9.7	479	17	US-10-739-930-9815	Sequence 9815, App
45	74.5	9.7	580	15	US-10-282-122A-73164	Sequence 73164, A

#### ALIGNMENTS

#### RESULT 1

US-10-293-239-17  
; Sequence 17, Application US/10293239  
; Publication No. US20030119043A1  
; GENERAL INFORMATION:  
; APPLICANT: Tanner, Stephan  
; APPLICANT: de la Chapelle, Albert  
; TITLE OF INVENTION: BAC expression as a diagnostic marker for acute leukemia  
; FILE REFERENCE: 22727/04101  
; CURRENT APPLICATION NUMBER: US/10/293,239  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: US 60/348,210  
; PRIOR FILING DATE: 2001-11-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 145  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-293-239-17

Query Match 84.6%; Score 649; DB 14; Length 145;

Best Local Similarity 83.4%; Pred. No. 1.1e-59;

Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MCGCGSRADAEPRYIESWTRETSTWLTYSDDALPSAAATDGGPAGGLHAGVLEGL 60

Db 1 MCGCGSRADAEPRYIESWTRETSTWLTYSDDALPSAAATDGGPAGGLHAGVLEGL 60

Qy 61 SSNGLVLPAPGGTANPEKKMCCGCPNSNLSGGPLTKONGWATEAKRDKMSAR 120

Db 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPSGSLTKONGLOTTEAKRDKMPAK 120

Qy 121 EVAINVTENIRQMDRSKRVTKNCIN 145

Db 121 EVINVTDSIQMDRSRRTKNCVN 145

```
RESULT 2
US-10-177-390-30
; Sequence 30, Application US/10177390
; Publication No. US20030143743A1
; GENERAL INFORMATION:
; APPLICANT: Schuler, Gerold
; APPLICANT: N.V. Anwerps Innovatiecentrum
; TITLE OF INVENTION: Improved transfection of Eucaryotic Cells with Linear
; FILE OF INVENTION: Polynucleotides by Electroporation
; FILE REFERENCE: 021505w/JH/ml
; CURRENT APPLICATION NUMBER: US/10/177,390
; CURRENT FILING DATE: 2002-06-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-390-30

Query Match      84.6%; Score 649; DB 14; Length 145;
Best Local Similarity 83.4%; Pred. No. 1.1e-59;
Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 60
D 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 60
QY 61 SSGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 120
D 61 PSNGVPRSTAGGIPNPEKKTNCETQCPNPQSLSSGGLTKQNGLWATEAKRDKRMSAR 120
QY 121 EVAVNTENIRQMDRSKRVTNKCIN 145
D 121 EVTVNTDSIQMDRSRRITKNCVN 145

RESULT 3
US-10-293-239-19
; Sequence 19, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-19

Query Match      81.0%; Score 621.5; DB 14; Length 180;
Best Local Similarity 67.2%; Pred. No. 1.1e-56;
Matches 121; Conservative 10; Mismatches 14; Indels 35; Gaps 1;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 53
D 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 60
QY 54 -----GVLEDGLSSNGVLRPAAPGCIANPEKKNCGT 85
D 61 KIKAPTDSVDEGLFSASKMAPLAVFSGHMGLEDGLPSNGVPRSTAGGIPNPEKKTNCET 120
QY 86 QCPNSQNLSSGGLTKQNGLWATE 145

RESULT 4
US-10-293-239-21
; Sequence 21, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-21

Query Match      61.6%; Score 472.5; DB 14; Length 149;
Best Local Similarity 65.3%; Pred. No. 3.4e-41;
Matches 94; Conservative 3; Mismatches 12; Indels 35; Gaps 1;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 53
D 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 60
QY 54 -----GVLEDGLSSNGVLRPAAPGCIANPEKKNCGT 85
D 61 KIKAPTDSVDEGLFSASKMAPLAVFSGHMGLEDGLPSNGVPRSTAGGIPNPEKKTNCET 120
QY 86 QCPNSQNLSSGGLTKQNGLWATE 109
D 121 QCPNPQSLSSGGLTKQNGLQTTTE 144

RESULT 5
US-10-293-239-20
; Sequence 20, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-20

Query Match      37.5%; Score 288; DB 14; Length 73;
Best Local Similarity 91.4%; Pred. No. 2.7e-22;
Matches 53; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 58
D 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 58

RESULT 6
US-10-293-239-20
; Sequence 20, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-20

Query Match      37.5%; Score 288; DB 14; Length 73;
Best Local Similarity 91.4%; Pred. No. 2.7e-22;
Matches 53; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 58
D 1 MCGGSRADAEIPRYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGL 58
```

```
US-10-293-239-18
; Sequence 18, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-18

Query Match      36.2%; Score 278; DB 14; Length 54;
Best Local Similarity 94.4%; Pred. No. 2e-21;
Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCGGSRADAIEPRYVESWTRETESTWLTYYTDSALPSSAAATDGPAGGLHAG 54
Db 1 MCGGSRADAIEPRYVESWTRETESTWLTYYTDSALPSSAAATDGPAGGLHAG 54

RESULT 7
US-10-293-239-22
; Sequence 22, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-22

Query Match      35.5%; Score 272; DB 14; Length 80;
Best Local Similarity 94.3%; Pred. No. 1.4e-20;
Matches 50; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCGGSRADAIEPRYVESWTRETESTWLTYYTDSALPSSAAATDGPAGGLHAG 53
Db 1 MCGGSRADAIEPRYVESWTRETESTWLTYYTDSALPSSAAATDGPAGGLHAG 53

RESULT 8
US-10-293-239-37
; Sequence 37, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-37

Query Match      18.3%; Score 140; DB 14; Length 25;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RADAIEPRYVESWTRETESTWLTYYT 31
Db 1 RADAIEPRYVESWTRETESTWLTYYT 25

RESULT 9
US-10-293-239-35
; Sequence 35, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-35

Query Match      12.9%; Score 99; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DAIEPRYVESWTRETEST 26
Db 1 DAIEPRYVESWTRETEST 18

RESULT 10
US-10-425-115-236244
; Sequence 236244, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 236244
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_147036C.1.pep
US-10-425-115-236244

Query Match      11.5%; Score 88; DB 17; Length 394;
Best Local Similarity 26.4%; Pred. No. 1.9;
Matches 39; Conservative 23; Mismatches 42; Indels 44; Gaps 9;
```

```

US-10-767-701-48905

Query Match          11.0%; Score 84; DB 16; Length 147;
Best Local Similarity 26.3%; Pred.No. 1.4;
Matches 30; Conservative 18; Mismatches 46; Indels 20; Gaps 5;

Qy      36 LPSAANTDSGPEA-----GGLHAGVLEDLSSNGVLRPAPAGGIANPEKKNCGTQCPNS 90
       ||| : | | | | | | | | | | | | | | | | : | | | | |
Db      29 LPAA LAPEDAAGEAGDNGGAHRG-----PHRQAEARAGAGA---RDGALHCGGPCPHS 79

Qy      91 QN-----LSGGPLTKQK-NGLWATEAKRDAKMSAREVAIVNTENTQMDRSKR 138
       ||| : | | | | | | | | | | | | | | | | : | | | | |
Db      80 DSCIGRWTSYGTSGHPKCSGLLVOLLROMPRSSAEQVRKEIKKIQEQEQRVRR 133

```

RESULT 13  
US-09-801-368-60  
Sequence 60, Application US/09801368  
Patent No. US20020128250A1  
GENERAL INFORMATION:  
APPLICANT: Busby, Robert  
APPLICANT: Cali, Brian  
APPLICANT: Hecht, Peter  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin  
APPLICANT: Maxon, Mary  
APPLICANT: Milne, Todd  
APPLICANT: No. US20020128250A1man, Thea  
APPLICANT: Royer, John  
APPLICANT: Salama, Sofie  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeff  
APPLICANT: Summers, Eric  
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
FILE REFERENCE: 109272.147  
CURRENT APPLICATION NUMBER: US/09/801.368

```

:
: PRIOR FILING DATE: 1999-10-20
:
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 60
: LENGTH: 1433
: TYPE: PrT
: ORGANISM: Saccharomyces cerevisiae
: US-09-801-368-60
:

```

```

QY 36 LPSAAANDSPEAGGLHAGVLEDGLSSGVLRPAAPGGIANPEKKMNCGTCCPNQNLSS 95
Db 927 LPPATTTSLLPLFLFGSQSKNSLENRQRPNNVKR-----ENPEHYIYGNDSNNNNNSEA 979
QY 96 G--PLTQKQGLWATEAKRAKRAKMSAREVAINVTENIR--QMDRSKRV 140
Db 980 GHSPMTNTTNGKELKYEKDAKR-NAKGGISKGENAHNFQNDTKNWS 1027

RESULT 14
US-10-149-310-112
; Sequence 112, Application US/10149310
; Publication No. US20040077039A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; APPLICANT: Madden, Kevin T.
; APPLICANT: Maxon, Mary
; APPLICANT: Sherman, Amir
; TITLE OF INVENTION: Modulation of Secondary Metabolite Production by
; TITLE OF INVENTION: Zinc Binuclear Cluster Proteins

```



Thu Nov 18 08:32:22 2004

```
; FILE REFERENCE: 14184-019US1
; CURRENT APPLICATION NUMBER: US/10/149,310
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: PCT/US01/29288
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 60/233,564
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 112
; LENGTH: 1433
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-149-310-112

Query Match      11.0%; Score 84; DB 15; Length 1433;
Best Local Similarity 29.4%; Pred.No.26;
Matches 32; Conservative 15; Mismatches 50; Indels 12; Gaps 4;

QY 36 LPSAAATDSGPEAGLHAGVLEDGLSSNGVLRPAAPGGIANPEKKMNCCTQCPNSQNLSS 95
Db 927 LPPATTTSLKPLFGSQSKNSLENRQRTPNVKR-----ENPEHEVLYGNDSDNNNNSEA 979

QY 96 G--PLTQKQGLWATEAKRDAKMSAREVAINVTENIR--QMDRSKRVY 140
Db 980 GHSPMTNTTNGNKRLLKYEKDAKR-NAKDGISKGNAHNFQNDTKKMS 1027

RESULT 15
US-10-425-115-275422
; Sequence 275422, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 275422
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_182776C.1.pep
US-10-425-115-275422

Query Match      10.8%; Score 82.5; DB 17; Length 264;
Best Local Similarity 34.4%; Pred.No.4.3;
Matches 31; Conservative 8; Mismatches 32; Indels 19; Gaps 4;

QY 32 DSDALPSAAATDSGPEAGLHAGVLEDGLSSNGVLRPAAPGGI---ANPEKKMNCCTQCP 88
Db 137 DSDDAFAAAAHDDQPAAVVAAAGGCGGSSSG-LPPAAGAATAEAEPSSLGLPLP 195

QY 89 -----NSQNLSSGPLTKONG 104
Db 196 AAEPAAEAADDSRNSQSQAS-PLLEEGEG 224

Search completed: November 17, 2004, 15:48:54
Job time : 82.3333 secs
```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 15:00:21 ; Search time 19.6667 Seconds  
(without alignments)  
709.395 Million cell updates/sec

Title: US-10-705-716A-8

Perfect score: 767

Sequence: 1 MCGGSRADAEPRYIESWT.....VTENIQMDRSKRVTKNCIN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 9621673 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_79.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87.5	11.4	778	2 A35970	erythrocyte-bindin
2	84	11.0	365	1 GNVSSC	genome polyprotein
3	84	11.0	1433	2 S54587	CAR8 protein - yea
4	79.5	10.4	571	2 T03456	hypothetical prote
5	79.5	10.4	1063	2 T03743	bifocal protein -
6	79	10.3	876	1 A57988	regulatory protein
7	78.5	10.2	600	2 C69899	conserved hypothet
8	78	10.2	269	2 T37073	hypothetical prote
9	77.5	10.1	848	2 G90128	elongation factor
10	77.5	10.1	2559	2 T09144	probable guanine n
11	77	10.0	781	2 T41551	hypothetical prote
12	76	9.9	516	2 JF0301	inulinase (EC 3.2.
13	76	9.9	645	2 S19156	serotonin receptor
14	74.5	9.7	601	2 AH0784	probable transport
15	74	9.6	1113	2 T14260	period protein fer
16	73.5	9.6	281	2 T02813	hypothetical prote
17	73.5	9.6	346	2 S61991	hypothetical prote
18	73.5	9.6	435	2 AG1028	prepilin (imported
19	72.5	9.5	451	2 S13337	tubulin alpha-2 ch
20	72.5	9.5	815	2 H69775	hypothetical prote
21	72.5	9.5	1050	2 S54640	KCS1 protein - yea
22	72	9.4	424	2 T01383	GTPase-activating
23	72	9.4	4957	2 T03455	ALR protein - huma
24	72	9.4	5262	2 T03454	ALR protein - huma
25	71.5	9.3	3488	2 T34418	hypothetical prote
26	71	9.3	652	2 G82962	hypothetical prote
27	71	9.3	1840	2 T30250	GPI protein - mous
28	71	9.3	5327	2 T13564	microtubule-associ
29	70.5	9.2	251	2 C75521	cytochrome-related

## ALIGNMENTS

## RESULT 1

A35970

erythrocyte-binding protein - Plasmodium knowlesi

C:Species: Plasmodium knowlesi

C:Date: 14-Dec-1990 #sequence\_revision 14-Dec-1990 #text\_change 09-Jul-2004

C:Accession: A35970

R:Adams, J.H.; Hudson, D.B.; Torii, M.; Ward, G.E.; Wellem, T.E.; Alkawa, M.; Miller, L

Cell 63, 141-153, 1990

A:Title: The Duffy receptor family of plasmodium knowlesi is located within the micronem

A:Reference number: A35970; PMID:91004213; PMID:2170017

A:Accession: A35970

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-778 <ADA>

A:Cross-references: UNIPROT:P22545; GB:M68518; GB:M37513; NID:g160273; PID:g160274

C:Keywords: transmembrane protein

Query Match 11.4%; Score 87.5; DB 2; Length 778;

Best Local Similarity 31.8%; Pred. No. 3.1;

Matches 28; Conservative 13; Mismatches 44; Indels 3; Gaps 3;

QY 31 TDSALPSAAATDSQPEAGLHAGVLEGLSSNGV-LRPAAPGGIANPEKKNKC-GTQCP 88

Db 352 TVSSDVPSVGGKDSQPSASHALAGENGVEHVTDTPEKDEKADPKDIEVKGQDT 411

QY 89 NSQNLSS-GPITQKQNGLMWATEAKRDAK 115

Db 412 DDRSQSLGPHPTDERATLTGETHMEKDT 439

## RESULT 2

GNVSSC

genome polyprotein - sugarcane mosaic virus (strain SC) (fragment)

N:Contains: carboxyl end of nuclear inclusion protein b; coat protein

C:Species: sugarcane mosaic virus, SCMV

C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 09-Jul-2004

C:Accession: PH0207

R:Frenkel, M.J.; Jilka, J.M.; McKern, N.M.; Strike, P.M.; Clark Jr., J.M.; Shukla, D.D.;

J. Gen. Virol. 72, 237-242, 1991

A:Title: Unexpected sequence diversity in the amino-terminal ends of the coat proteins o

A:Reference number: PH0207; PMID:91132116; PMID:1993866

A:Accession: PH0207

A:Molecule type: genomic RNA

A:Residues: 1-365 <PRE>

A:Cross-references: UNIPROT:P25242; GB:D00948; NID:g222123; PIDN:BAA00796.1; PID:g222124

C:Superfamily: tobacco etch virus genome polyprotein

C:Keywords: coat protein; inclusion protein

F:1-52/Product: nuclear inclusion protein b (fragment) #status predicted <IPB>

F:53-365/Product: coat protein #status predicted <COA>

Query Match 11.0%; Score 84; DB 1; Length 365;

Best Local Similarity 21.6%; Pred. No. 2.7;

```

C;Accession: T43456
R;Foustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z22516
A;Accession: T43456
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-571 <BA>
A;Cross-references: UNIPROT:O75175; EMBL:AL133647
A;Experimental source: adult testis; clone DKFZp434L061
C;Genetics:
A;Note: DKFZp434L061.1
C;Superfamily: hydroxyproline-rich glycoprotein

Query Match      10.4%; Score 79.5; DB 2; Length 571;
Best Local Similarity 30.9%; Pred. No. 12;
Matches 29; Conservative 6; Mismatches 34; Indels 25; Gaps 3

Qy 31 TSDALPSAAATDSGPAGGLH-----AGVLEDGLSSNGVLRPAAP-- 71
Db 111 TDSEVSQSPAKNGSKPVHSNQHPQSPAVPTYPGPPPAASALSTTPGNGVGPAPAAPPS 170

Qy 72 --GGIANPEKKMNCGTQCPNSQNLS-----SQPLT 99
Db 171 ALGPKASPAPSHNSGTPAPYAQAAPAPSPGST 204

RESULT 5
T03743
bifocal protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C;Accession: T03743
R;Bahri, S.M.; Yang, X.Y.; Chia, W.
Mol. Cell. Biol. 17, 5521-5529, 1997
A;Title: The Drosophila bifocal gene encodes a novel protein which colocalizes
A;Reference number: Z15048; MUID:97415628; PMID:9271427
A;Accession: T03743
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1063 <BA>
A;Cross-references: UNIPROT:O16125; EMBL:AF011791; NID:g2388667; PIDN:AAB699951
C;Genetics:
A;Cross-references: FlyBase:FBgn0014133
A;Note: bifocal

Query Match      10.4%; Score 79.5; DB 2; Length 1063;
Best Local Similarity 20.7%; Pred. No. 25;
Matches 39; Conservative 24; Mismatches 58; Indels 67; Gaps 6

Qy 5 GSRADALIEPRVYESWTR-----ETESTWLTITDSDAL-----PSAAAT- 42
Db 83 GAIADFTPATISSTQKRNMIGSEKSEKISINTNSDSTGGHSHVAVSLSPDAAATT 142

Qy 43 -----DSGPEAGGLHAGVLEDGLSSNGVLRPAAPG 72
Db 143 NVTVTPDKQRSSLLNTRSQREVMRYILSSGSRDGELESGEQAGVYSNRCGEVETG 202

Qy 73 GIANPEKKMNCGTQCPNSQNLSGSLTQKQGLMATEAKRDARMSAREVAINVTENIQ 132
Db 203 TIGSPSSSAN--QNPENHLK---TKCKPQSVAEKPSAKE-----TIVDNSK 247

Qy 133 MDRSKRVT 140
Db 248 CSKTKSIS 255

RESULT 6
A57988
regulatory protein area - Emicella nidulans
C;Species: Emicella nidulans, Aspergillus nidulans
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C;Accession: A57988; S10017; S70168; S72883

```

R;Kudla, B.; Caddick, M.X.; Langdon, T.; Martinez-Rossi, N.M.; Bennett, C.F.; Sibley, S.; EMBL J. 9, 1355-1364, 1990  
A:Title: The regulatory gene areA mediating nitrogen metabolite repression in Aspergillus  
A:Reference number: S10017; MUID:90228331; PMID:1970293  
A:Accession: A57988  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-876 <KUD>  
A:Cross-references: UNIPROT:P17429; EMBL:X52491; NID:g1019911; PIDN:CAA36731.1; PID:g115  
A:Note: this sequence represents reinterpretation to include two exons  
A:Accession: S10017  
A:Molecule type: DNA  
A:Residues: 158-876 <KU2>  
A:Cross-references: EMBL:X52491  
A:Note: this sequence represents the authors' original translation  
R;Langdon, T.; Sheerins, A.; Ravagnani, A.; Gielkens, M.; Caddick, M.X.; Arst Jr., H.N.  
Mol. Microbiol. 17, 877-888, 1995  
A:Title: Mutational analysis reveals dispensability of the N-terminal region of the Aspe  
A:Reference number: S70167; MUID:96123430; PMID:8596437  
A:Accession: S70168  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-791, 'A', 793, 'T', 795, 'SPGTNS', 802-876 <LAN>  
A:Cross-references: EMBL:X52491  
R;Caddick, M.X.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: S72883  
A:Accession: S72883  
A:Molecule type: DNA  
A:Residues: 1-876 <CAD>  
A:Cross-references: EMBL:X52491; NID:g1019911; PIDN:CAA36731.1; PID:g1154625  
C:Genetics:  
A:Gene: areA  
A:Introns: 147/2  
C:Function:  
A:Description: mediates nitrogen metabolite repression  
C:Superfamily: nitrogen regulatory protein nit-2; GATA-type zinc finger homology  
C:Keywords: DNA binding; transcription regulation; zinc finger  
F:670-723/Domain: GATA-type zinc finger homology <GZF>  
F:673-697/Region: zinc finger GATA motif

Query Match 10.3%; Score 79; DB 1; Length 876;  
Best Local Similarity 28.4%; Pred. No. 22;  
Matches 31; Conservative 10; Mismatches 28; Indels 40; Gaps 5;  
QY 57 EDGLSSNGVLRPAAPGGIANPEKK-----MNCGTQC-----PNSQLSSG----- 96  
DB 645 ESLNSAAPSAPSPGCTGNGEQNGPTCTCTTTLWRNPGQPLCNACGLFLKLH 704  
QY 97 ----PLTQKQGLWATEAKRDARMSAREVAINVTENIROMDRSKRVTK 141  
DB 705 GWRPLSLK-----TDVKKRNNSANSLAVG-----SSRVSK 737

RESULT 7  
C69899  
conserved hypothetical protein yobL - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: C69899  
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
A.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl T.M.; Portetel  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, F.; Sekowska, A.; Ser  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A9580; MUID:98044033; PMID:9384377  
A:Accession: C69899  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-600 <KUN>  
A:Cross-references: UNIPROT:O34330; GB:Z99114; GB:AL009126; NID:g2634230; PIDN:CAB13792.  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yobL

Query Match 10.2%; Score 78; DB 2; Length 600;  
Best Local Similarity 23.7%; Pred. No. 16;  
Matches 40; Conservative 22; Mismatches 56; Indels 51; Gaps 7;  
QY 9 DAIEPRYVESWTRE-----TESTWLTYSDDALPSAAATDSGPAGGLHAGVLEDGLS 61  
DB 342 DAISAAIEESYQKQMVNGDAYSRSRWTV-----AIGSVAVAVVGTGK-----AGAINKADA 393  
QY 62 SNGVLRPAAPGGIA-----NPEKMNCTQCP-----NSQNLSSGLT----- 99  
DB 394 AGKINKASQAGKKIKVDKIPDLFPNPKYKIALADNPYNNVDSQNLKNEILLTNAKIP 453  
QY 100 -----QKQGLWATEAKRDARMSAREVAINVTENIROMDRSKRV 139  
DB 454 DGTXPFTGQKSPPLNKEKYD-----AYEIEGKVKAKGKVDVSRVV 497

RESULT 8  
T37073  
hypothetical protein SCJ30.06c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37073  
R;Sanders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21621  
A:Accession: T37073  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-269 <SAN>  
A:Cross-references: UNIPROT:Q9S1Y3; EMBL:AL109973; PIDN:CAB53301.1; GSPDB:GN00070; SCOE  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOE08B:SCJ30.06c

Query Match 10.2%; Score 78; DB 2; Length 269;  
Best Local Similarity 25.8%; Pred. No. 6.9;  
Matches 25; Conservative 10; Mismatches 38; Indels 24; Gaps 4;  
QY 11 IEPRYVESWT-----RETESTWLTYSDDALPSAAATDSGPAGGLHAGVLEDGL 60  
DB 176 LPPTHVEQVFPVNFHCRPRRVLRSGLTFSPFTVLPSRGCCSGERTGG----- 225  
QY 61 SNGVLRPAAPGGIANPEKMNCG--TQCPNSQNLSS 95  
DB 226 -GGGVVTGAGGIV-PEACRCGRAAHAPTNASMRS 260

RESULT 9  
G90128  
elongation factor EF-2 [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: G90128  
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei  
Nature 410, 1091-1096, 2001  
A:Title: The highly reduced genome of an enslaved algal nucleus.  
A:Reference number: A99082; MUID:11323671; PMID:11323671  
A:Accession: G90128  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-848 <DOU>  
A:Cross-references: UNIPROT:Q98S60; GB:AF080301; NID:gl3794345; PIDN:AAK39722.1; GSPDB:G  
C:Genetics:  
A:Gene: EF2  
A:Map position: 3  
A:Genome: nucleomorph  
C:Superfamily: translation elongation factor 2; translation elongation factor Tu homolog  
C:Keywords: nucleomorph

Query Match 10.1%; Score 77.5; DB 2; Length 848;  
Best Local Similarity 25.2%; Pred. No. 30;  
Matches 38; Conservative 23; Mismatches 65; Indels 25; Gaps 7;

QY 10 ATEPRYEWTRTESTWLTYSDDLPSAAATDSGE---AGGLHAGV---LEDGLS 61  
DB 493 AVEPKNPDLFKLIEGL-KRLSKSDPLVQCMWTESEGEHIVAGAEHLLEICLKDLQDDFM 551

QY 62 SNGVLRPAAP-----GGIANPEKKNCGTQCPNSN---LSSGFLTQKQNGWLWATEA 110  
DB 552 NGAELKISQPIVSGYRETVGVPINPETAVCLSKSPKHNRIYCYAPLPE---GL--SEA 606

QY 111 KRPAKMGAREVAINVTENIRQMDRSKRVTK 141  
DB 607 IDGKIKSDPEKTRIKELKQKFDMEEDIK 637

RESULT 10  
T09144  
probable guanine nucleotide exchange factor RhoGEF2 - fruit fly (Drosophila melanogaster)  
N;Alternate names: Shar pei/DRHOGF2  
C:Species: Drosophila melanogaster  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: T09144; T09223  
R;Haacker, U.; Perrimon, N.  
submitted to the EMBL Data Library, October 1997  
A:Reference number: Z16586  
A:Accession: T09144  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2559 <HAE>  
A:Cross-references: UNIPROT:O44113; EMBL:AF031930; NID:G2687355; PID:G2687356  
R;Barrett, K.; Leptin, M.; Settlemann, J.  
Cell 91, 905-915, 1997  
A:Title: The Rho GTPase and a putative RhoGEF mediate a signaling pathway for the cell s  
A:Reference number: Z16618; MUID:98088790; PMID:9428514  
A:Accession: T09223  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-676,'L',678-837,'L',839-889,'S',891-1243,'D',1245-1358,'E',1360-1368,'R',1  
A:Cross-references: EMBL:AF032870; NID:G2760367; PIDN:AAK38820.1; PID:G2760368  
C:Genetics:  
A:Gene: rhoGEF2  
A:Cross-references: FlyBase:FBgn0023172  
A:Map position: 2; 53F1-2  
A:Note: orchestrates cell shape changes during gastrulation  
C:Function:  
A:Description: mediates actin rearrangements required for cell shape changes during gas  
C:Superfamily: rat Munc13-3 protein; protein kinase C zinc-binding repeat homolog  
C:Keywords: embryo; GTP exchange; signal transduction  
F;1151-1200/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 10.1%; Score 77.5; DB 2; Length 2559;  
Best Local Similarity 23.9%; Pred. No. 1.1e+02;  
Matches 28; Conservative 16; Mismatches 56; Indels 7; Gaps 2;

QY 5 GSRADAIEPRYEWTRTESTWLTYSDDLPSAAATDSGEAGGLHAGVLEGLSSNG 64  
DB 799 GSSPDNMHPRPDRITTKTSGW-EIVEKDGESSPPGTPPPPYLLSSSHMTVLEDPENNR 857

QY 65 VLRAAAGGIANPEKKNCGTQCPNSQLSSGFLTQKQNGWLWATEAKDKRMSARE 121  
DB 858 GAAAGGCVFTIESHQ-----FTPWAGSSPIPLSHSNMHAQAQNDTQKEIISME 908

# RESULT 11

T41551

hypothetical protein SPCC70.05c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T41551

R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.

submitted to the EMBL Data Library, June 1998

A:Reference number: Z22001

A:Accession: T41551

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-781 <WOO>

A:Cross-references: UNIPROT:O74526; EMBL:AL023794; PIDN:CAA19355.1; GSPDB:GN00068; SPDB

A:Experimental source: strain 972h-; cosmid C70

C:Genetics:

A:Gene: SPDB:SPCC70.05c

A:Map position: 3

Query Match 10.0%; Score 77; DB 2; Length 781;

Best Local Similarity 26.6%; Pred. No. 30;

Matches 34; Conservative 16; Mismatches 64; Indels 14; Gaps 5;

QY 16 YESWTRETESTWLTYSDDLPSAAATDS-----GPEAGGLHAGVLEGLSSNGV 65

DB 326 YESWPHSTEFDMFTYAVSGSLKLTPOGTGDCINPANPEFSPGYSGSKMSKSDDNVGSAN 385

QY 66 LRPAAPGGIANPEKKNCG-TQCP--NSQNLSSGFLTQKQNGWLWATEAKDKRMSAREV 122

DB 386 TAPNSNTSANSSEGNQNGTTPYIPKPTNISEIPKPKL-SGFIPIPPYAKKVVPLSAKYK 444

QY 123 AINVTENI 130

DB 445 LVDETQDM 452

# RESULT 12

JE0301

inulinase (EC 3.2.1.7) - Aspergillus niger

C:Species: Aspergillus niger

C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 09-Jul-2004

C:Accession: JE0301

R;Ohta, K.; Akimoto, H.; Matsuda, S.; Toshimitsu, D.; Nakamura, T.

BioSci. Biotechnol. Biochem. 62, 1731-1738, 1998

A:Title: Molecular cloning and sequence analysis of two endoinulinase genes from Asperg

A:Reference number: JE0301; MUID:99022189; PMID:9805373

A:Accession: JE0301

A:Molecule type: mRNA

A:Residues: 1-516 <OHT>

A:Cross-references: UNIPROT:O74641; UNIPROT:Q8X217; UNIPROT:Q96WZ8; UNIPROT:O74642; DDB

C:Superfamily: Penicillium purporogenun inulinase

C:Keywords: glycosidase; hydrolase

Query Match 9.9%; Score 76; DB 2; Length 516;

Best Local Similarity 42.3%; Pred. No. 23;

Matches 22; Conservative 6; Mismatches 16; Indels 8; Gaps 2;

QY 23 TESTWLTYSDDLPSAAATDSG-----PEAGGLHAGVLEGLSSNGVL 66

DB 413 SEQIVINTOSNATLSVDRTESGDISYDPAAGGVHTAKLEEDGTGLSVIRVL 464

# RESULT 13

S19156

serotonin receptor 2B - fruit fly (Drosophila melanogaster)

N;Alternate names: 5-hydroxytryptamine receptor 2B (5-HTR2B)

C:Species: Drosophila melanogaster

C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 09-Jul-2004

C:Accession: S19156; S19154

R;Saudou, F.; Boscher, U.; Amlaiky, N.; Plassat, J.L.; Hen, R.

EMBO J. 11, 7-17, 1992







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 14:58:39 ; Search time 98 Seconds  
(without alignments)  
530.773 Million cell updates/sec

Title: US-10-705-716a-2

Perfect score: 767

Sequence: 1 MGCGSRADAEIPRYESWT.....VTENIQMDRSKRVTKNCIN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description
1	767	100.0	145	8	ADO48473	ADO48473	Rat PTH r
2	746	97.3	145	8	ADO48479	ADO48479	Mouse PTH
3	645	84.1	145	4	AAB95018	AAB95018	Human pro
4	645	84.1	145	5	AAO19498	AAO19498	HSL prote
5	645	84.1	145	6	ABR58646	ABR58646	Human can
6	645	84.1	145	7	ADC31800	ADC31800	Human nov
7	645	84.1	145	7	ADM46959	ADM46959	Brain and
8	645	84.1	145	8	ADO48475	ADO48475	Human PTH
9	617.5	80.5	180	7	ADM46961	ADM46961	Brain and
10	473.5	61.7	149	7	ADM46963	ADM46963	Brain and
11	335.5	43.7	92	5	ADO81902	ADO81902	Human dio
12	294	38.3	54	8	ADO48481	ADO48481	Mouse PTH
13	288	37.5	73	7	ADM46962	ADM46962	Brain and
14	278	36.2	54	7	ADM46960	ADM46960	Brain and
15	278	36.2	54	8	ADO48477	ADO48477	Human PTH
16	272	35.5	80	7	ADM46964	ADM46964	Brain and
17	140	18.3	25	7	ADM46979	ADM46979	Brain and
18	99	12.9	18	7	ADM46977	ADM46977	Brain and
19	92.5	12.1	778	2	AAR13456	AAR13456	Duffy rec
20	89	11.6	16	8	ADO48482	ADO48482	PTH respo
21	87.5	11.4	260	4	ABG09899	ABG09899	Novel hum
22	87.5	11.4	592	4	ABG15607	ABG15607	Novel hum
23	84	11.0	718	6	ABU17344	ABU17344	Protein e
24	84	11.0	726	6	ADA36828	ADA36828	Actinotoba
25	83.5	10.9	700	5	AAE25052	AAE25052	Stenotrop

ALIGNMENTS

RESULT 1

ADO48473  
ID ADO48473 standard; protein; 145 AA.

XX AC ADO48473;

XX DT 12-AUG-2004 (first entry)

XX DE Rat PTH responsive gene protein.

XX KW PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation; transgenic animal; osteopathic; gene therapy; osteoporosis; rat.

XX OS Rattus sp.

XX PN WO2004044152-A2.

XX PD 27-MAY-2004.

XX PF 10-NOV-2003; 2003WO-US035655.

XX PR 12-NOV-2002; 2002US-0425532P.

XX (AMHP ) WYETH.

XX Robinson JA, Stojanovic-Susulic V, Babić P, Murrills RJ;

XX WPI; 2004-420299/39.

XX N-ESDB; ADO48472.

XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in preparing a composition for diagnosing, treating or preventing bone related disorders, e.g., osteoporosis.

XX Claim 9; SEQ ID NO 2; 169pp; English.

XX The invention relates to a novel PTH responsive gene (PAIGB) fragment encoding a polypeptide. The invention further comprises: a chimeric construct comprising the isolated nucleic acid fragment operatively linked to suitable regulatory sequences; a host cell transformed with the chimeric construct; a vector comprising the nucleic acid fragment; obtaining a nucleic acid fragment encoding the polypeptide; a method for obtaining a polypeptide; detecting the presence of the nucleic acid fragment; an antibody that specifically binds to one or more epitopes of a PAIGB polypeptide; a composition for regulating bone-forming activity in a mammal comprising the nucleic acid fragment, polypeptide or antibody; an agent that alters the expression of PAIGB gene or polypeptide; determining whether an agent alters the expression of PAIGB mRNA;

Abp35624 Fungal ZB  
Aae14866 S. Clavul  
Adj70139 Human hea  
Aam40296 Human pol  
Aam40295 Human pol  
Aam42081 Human pol  
Aam42082 Human pol  
Aau31506 Novel hum  
Abg21379 Novel hum  
Abg15431 Novel hum  
Aag94930 Shrimp wh  
Aay00097 Enterococ  
Abp43316 E faecali  
Abu88344 E. faecal  
Abu13595 Enterococ  
Aay00096 Enterococ  
Abp43315 E faecali  
Abu88343 E. faecal  
Abu13594 Enterococ  
Abj37782 Human tum

26 83 10.8 1433 5 ABP35624  
27 82 10.7 339 6 AAE14866  
28 81.5 10.6 600 7 ADJ70139  
29 81.5 10.6 852 4 AAM40296  
30 81.5 10.6 872 4 AAM40295  
31 81.5 10.6 886 4 AAM42081  
32 81.5 10.6 886 4 AAM42082  
33 81.5 10.6 974 4 AAU31506  
34 80.5 10.5 275 4 ABG21379  
35 80.5 10.5 275 4 ABG15431  
36 80.5 10.5 1100 4 AAG94930  
37 79.5 10.4 286 2 AAY00097  
38 79.5 10.4 286 5 ABP43316  
39 79.5 10.4 286 6 ABU88344  
40 79.5 10.4 286 6 ABU13595  
41 79.5 10.4 305 2 AAY00096  
42 79.5 10.4 305 5 ABP43315  
43 79.5 10.4 305 6 ABU88343  
44 79.5 10.4 305 6 ABU13594  
45 79.5 10.4 661 6 ABJ37782

CC screening agents for effectiveness in altering expression of the nucleic  
CC acid fragment; screening for agents useful for treating bone related  
CC disorders; evaluating the efficacy of a treatment of a bone related  
CC disorder in a subject; identifying polypeptides capable of binding to  
CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
CC related agent; a transgenic animal comprising the DNA; an animal model  
CC for the study of bone density modulation comprising a first group of  
CC animals composed of the transgenic animal and a second group of control  
CC animals; studying bone mass determinants; studying the modulation of bone  
CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
CC for treating bone related disorders; identifying whether an agent which  
CC has bone forming activity; and a stably transfected cell line comprising  
CC two constructs, the first construct comprising a ligand binding domain  
CC linked to a DNA binding domain which is linked to an activation domain  
CC all of which expression is driven by a constitutive promoter, the second  
CC construct comprising multiple copies of DNA binding elements linked to a  
CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
CC used to treat disorders by gene therapy. The nucleic acid is useful in  
CC preparing a composition for diagnosing, treating or preventing bone  
CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
CC responsive gene protein of the invention.  
XX  
SQ

Query Match 100.0%; Score 767; DB 8; Length 145;  
Best Local Similarity 100.0%; Pred. No. 3.9e-76;  
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSDALPSAAATDSGPPAGGLHAGVLEDP 60  
DB 1 MCGGSRADAIEPRYYESWTRETESTWLTYSDALPSAAATDSGPPAGGLHAGVLEDP 60  
QY 61 SSSGVLRLPAPGGIANPEKKMNCGTQCPNSQSLSSGGLTKQKGLWTEAKRDKRMSAR 120  
DB 61 SSSGVLRLPAPGGIANPEKKMNCGTQCPNSQSLSSGGLTKQKGLWTEAKRDKRMSAR 120  
QY 121 EVAISVTENIRQMDRSKRVTNCIN 145  
DB 121 EVAISVTENIRQMDRSKRVTNCIN 145

RESULT 2  
ADO48479  
ID ADO48479 standard; protein; 145 AA.  
XX  
AC ADO48479;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE Mouse PTH responsive gene protein.  
XX  
KW PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
KW transgenic animal; osteopathic; gene therapy; osteoporosis; mouse;  
KW murine.  
XX  
OS Mus sp.  
XX  
FN WO200404152-A2.  
XX  
PD 27-MAY-2004.  
XX  
PF 10-NOV-2003; 2003WO-US035655.  
XX  
PR 12-NOV-2002; 2002US-0425532P.  
XX  
PA (AMEP) WYETH.  
XX  
PI Robinson JA, Stojanovic-Susulic V, Babić P, Murrills RJ;  
XX  
DR WPI; 2004-420299/39.  
DR N-PSDB; ADO48478.  
DR

XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
PT preparing a composition for diagnosing, treating or preventing bone  
PT related disorders, e.g., osteoporosis.  
XX  
PS Claim 9; SEQ ID NO 8; 169pp; English.  
XX  
XX The invention relates to a novel PTH responsive gene (PAIGB) fragment  
CC encoding a polypeptide. The invention further comprises: a chimeric  
CC construct comprising the isolated nucleic acid fragment operatively  
CC linked to suitable regulatory sequences; a host cell transformed with the  
CC chimeric construct; a vector comprising the nucleic acid fragment;  
CC obtaining a nucleic acid fragment encoding the polypeptide; a method for  
CC obtaining a polypeptide; detecting the presence of the nucleic acid  
CC fragment; an antibody that specifically binds to one or more epitopes of  
CC a PAIGB polypeptide; a composition for regulating bone-forming activity  
CC in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
CC ; an agent that alters the expression of PAIGB gene or polypeptide;  
CC determining whether an agent alters the expression of PAIGB mRNA;  
CC screening agents for effectiveness in altering expression of the nucleic  
CC acid fragment; screening for agents useful for treating bone related  
CC disorders; evaluating the efficacy of a treatment of a bone related  
CC disorder in a subject; identifying polypeptides capable of binding to  
CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
CC related agent; a transgenic animal comprising the DNA; an animal model  
CC for the study of bone density modulation comprising a first group of  
CC animals composed of the transgenic animal and a second group of control  
CC animals; studying bone mass determinants; studying the modulation of bone  
CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
CC for treating bone related disorders; identifying whether an agent which  
CC has bone forming activity; and a stably transfected cell line comprising  
CC two constructs, the first construct comprising a ligand binding domain  
CC linked to a DNA binding domain which is linked to an activation domain  
CC all of which expression is driven by a constitutive promoter, the second  
CC construct comprising multiple copies of DNA binding elements linked to a  
CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
CC used to treat disorders by gene therapy. The nucleic acid is useful in  
CC preparing a composition for diagnosing, treating or preventing bone  
CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
CC responsive gene protein of the invention.  
XX  
SQ

Sequence 145 AA;  
Query Match 97.3%; Score 746; DB 8; Length 145;  
Best Local Similarity 97.2%; Pred. No. 8e-74;  
Matches 141; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSDALPSAAATDSGPPAGGLHAGVLEDP 60  
DB 1 MCGGSRADAIEPRYYESWTRETESTWLTYSDALPSAAATDSGPPAGGLHAGVLEDP 60  
QY 61 SSSGVLRLPAPGGIANPEKKMNCGTQCPNSQSLSSGGLTKQKGLWTEAKRDKRMSAR 120  
DB 61 SSSGVLRLPAPGGIANPEKKMNCGTQCPNSQSLSSGGLTKQKGLWTEAKRDKRMSAR 120  
QY 121 EVAISVTENIRQMDRSKRVTNCIN 145  
DB 121 EVAISVTENIRQMDRSKRVTNCIN 145  
RESULT 3  
AAB95018  
ID AAB95018 standard; protein; 145 AA.  
XX  
AC AAB95018;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:16726.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

```

XX OS Homo sapiens.
XX PN EP1074617-A2.
XX XX
XX PD 07-FEB-2001.
XX PF
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300255.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX XX
XX PA (HELI-) HELIX RES INST.
XX XX
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX XX
XX DR WPI; 2001-318749/34.
XX PT
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX PT length cDNAs defined in the specification, and for the detection and/or
XX PT diagnosis of the abnormality of the proteins encoded by the full-length
XX PT cDNAs.
XX PS
XX PS Claim 8; SEQ ID NO 16726; 2537pp + Sequence Listing; English.
XX CC
XX CC The present invention describes primer sets for synthesizing 5602 full-
XX CC length cDNAs defined in the specification. Where a primer set comprises:
XX CC (a) an oligo-dT primer and an oligonucleotide complementary to the
XX CC complementary strand of a polynucleotide which comprises one of the 5602
XX CC nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in the
XX CC specification. The primer sets can be used in antisense therapy and in
XX CC gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
XX CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX CC present invention
XX XX
XX SQ Sequence 145 AA;
Query Match 84.1%; Score 645; DB 4; Length 145;
Best Local Similarity 83.4%; Pred. No. 1.1e-62;
Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;
QY 1 MCGCGSRADALEPRYIESWTRETESTWLTYSALPSAAATDSGPAGGLHAGVLEDP 60
Db 1 MCGCGSRADALEPRYIESWTRETESTWLTYSALPSAAATDSGPAGGLHAGVLEDP 60
QY 61 SNGVLRPAAPGGIANPEKKNCTQCPNOSLSGGPLTQKQGLWTTTEAKRDKRMSAR 120
Db 61 PSNGVPRSTAGPGIPNPEKKNCTQCPNOSLSGGPLTQKQGLWTTTEAKRDKRMPAK 120
QY 121 EVAISVTENIRQMDRSKRVTNKCIN 145
Db 121 EVTINVTDISIQQMDRSRRITKNVCN 145
QY 121 EVAISVTENIRQMDRSKRVTNKCIN 145
Db 121 EVTINVTDISIQQMDRSRRITKNVCN 145
RESULT 4
AAO19498

```

```

ID AAO19498 standard; protein; 145 AA.
XX AC
XX AC AAO19498;
XX DT 20-DEC-2002 (first entry)
XX DE
XX DE HSI1 protein variant.
XX KW HSI1; variant; cancer; tumour; unigene cluster; cytostatic; metastasis;
XX KW EST; expressed sequence tag; colon cancer; stomach cancer; breast cancer;
XX OS Unidentified.
XX PN DE10103694-A1.
XX PD
XX PD 01-AUG-2002.
XX PF 26-JAN-2001; 2001DE-01003694.
XX PR 26-JAN-2001; 2001DE-01003694.
XX PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.
XX PI Brett D, Kemmner W;
XX DR WPI; 2002-644836/70.
XX DR N-PSDB; AAL50100.
XX PT
XX PT Diagnosis and therapy of tumors, by determining expression rates of
XX PT specific expressed sequence tags of the unigene cluster, and subsequently
XX PS blocking their expression.
XX PS Claim 10; Page 5; 10pp; German.
XX CC
XX CC The present invention relates to the use of expressed sequence tags
XX CC (ESTs), or variants, of the unigene cluster HSI69395 (HS1), HSI27144
XX CC (HS2) and/or HSI132793 (HS3) for diagnosis and therapy of tumours, in
XX CC which their expression rates in tumour cells and/or lymph nodes are
XX CC determined. The EST sequences are useful as prognostic markers of
XX CC survival of cancer patients (high levels of EST-related mRNA are
XX CC associated with a poor prognosis, specifically correlated with
XX CC development of metastases); and for diagnosis and/or therapy of solid
XX CC tumours, particularly of colon, stomach and breast. The present sequence
XX CC is a variant of the HSI protein shown in the exemplification of the
XX CC invention
XX SQ Sequence 145 AA;
Query Match 84.1%; Score 645; DB 5; Length 145;
Best Local Similarity 83.4%; Pred. No. 1.1e-62;
Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;
QY 1 MCGCGSRADALEPRYIESWTRETESTWLTYSALPSAAATDSGPAGGLHAGVLEDP 60
Db 1 MCGCGSRADALEPRYIESWTRETESTWLTYSALPSAAATDSGPAGGLHAGVLEDP 60
QY 61 SNGVLRPAAPGGIANPEKKNCTQCPNOSLSGGPLTQKQGLWTTTEAKRDKRMSAR 120
Db 61 PSNGVPRSTAGPGIPNPEKKNCTQCPNOSLSGGPLTQKQGLWTTTEAKRDKRMPAK 120
QY 121 EVAISVTENIRQMDRSKRVTNKCIN 145
Db 121 EVTINVTDISIQQMDRSRRITKNVCN 145
RESULT 5
ABR58646
ID ABR58646 standard; protein; 145 AA.
XX AC
XX AC ABR58646;
XX DT 09-JUL-2003 (first entry)

```

XX DE Human cancer related protein SEQ ID NO:303.  
 XX DE  
 XX DE Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;  
 XX DE heart disease; atherosclerosis; endometriosis.  
 XX DE  
 XX OS Homo sapiens.  
 XX OS  
 XX PN WO2003025138-A2.  
 XX PN  
 XX PD 27-MAR-2003.  
 XX PD  
 XX PF 17-SEP-2002; 2002WO-US029560.  
 XX PF  
 XX PR 17-SEP-2001; 2001US-0323469P.  
 XX PR  
 XX PR 20-SEP-2001; 2001US-0323887P.  
 XX PR  
 XX PR 13-NOV-2001; 2001US-0350666P.  
 XX PR  
 XX PR 08-FEB-2002; 2002US-0355145P.  
 XX PR  
 XX PR 08-FEB-2002; 2002US-0355257P.  
 XX PR  
 XX PR 12-APR-2002; 2002US-0372246P.  
 XX PR  
 XX PA (E0SB-) EOS BIOTECHNOLOGY INC.  
 XX PA  
 XX PI Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;  
 XX PI Zlotnik A;  
 XX PI  
 XX PR WPI; 2003-354600/33.  
 XX PR  
 XX PR N-PSDB; ACC72796.  
 XX PR  
 XX PT New genes that are up-regulated or down-regulated in cancers, useful as  
 XX PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as  
 XX PT therapeutic targets for screening drugs for treating these diseases.  
 XX PT  
 XX PS Claim 12; Page 753; 767pp; English.  
 XX PS  
 XX CC The present invention describes an isolated nucleic acid molecule, which  
 XX CC comprises the sequence of any of the genes that are up-regulated or down-  
 XX CC regulated in specific cancers (e.g. about 1031 genes up-regulated in  
 XX CC acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer  
 XX CC related gene nucleotide sequences which encode the proteins given in  
 XX CC AR5821 to AR58799. Also described: (1) determining the presence or  
 XX CC absence of a pathological cell in a patient; (2) an expression vector  
 XX CC comprising a nucleic acid molecule described above; (3) a host cell  
 XX CC comprising the vector; (4) an isolated polypeptide, which is encoded by  
 XX CC the nucleic acid; (5) an antibody that specifically binds the polypeptide  
 XX CC of (4); (6) specifically targeting a compound to a pathological cell in a  
 XX CC patient by administering to the patient the antibody above; and (7) a  
 XX CC drug screening assay. The nucleic acid is useful as diagnostic markers or  
 XX CC therapeutic targets. In particular, the nucleic acid is useful for  
 XX CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,  
 XX CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,  
 XX CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,  
 XX CC atherosclerosis and endometriosis. The nucleic acid is also useful in  
 XX CC drug screening, particularly for identifying agents for treating these  
 XX CC pathologies  
 XX CC  
 XX SQ Sequence 145 AA;  
 XX SQ  
 XX Query Match 84.1%; Score 645; DB 6; Length 145;  
 XX Best Local Similarity 83.4%; Pred. No. 1.1e-62;  
 XX Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MCGCGSRADALEPRYSWTRTETSTWLTVDLSAATDSGPGAGGLHAGVLEDGP 60  
 Db 1 MCGCGSRADALEPRYSWTRTETSTWLTVDLSAATDSGPGAGGLHAGVLEDGP 60  
 QY 61 SSGVLRPAAPGGIANPKKNNCTQCPNSQSLSSGPIITQKQGLWTEAKRDKMSAR 120  
 Db 61 PSNGVPRSTARPGIPNPKKNTCTQCPNPQSLSSGPIITQKQGLWTEAKRDKMSAR 120  
 QY 121 EVASIVTENIQRDMSRKVTNKN 145  
 Db 121 EVTINVTDSIQDMSRRITNKN 145

RESULT 6  
 ADC31800  
 ID ADC31800 standard; protein; 145 AA.  
 XX  
 AC ADC31800;  
 XX  
 XX 18-DEC-2003 (first entry)  
 XX  
 DE Human novel polypeptide sequence, SEQ ID NO:1882.  
 XX  
 XX Human; diagnostic; drug screening; forensics; gene mapping;  
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
 KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
 KW ulcers; osteoporosis; autoimmune disease; cancer;  
 KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
 KW neuroprotective; antianemic; anticoagulant; thrombolytic; vulnary;  
 KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
 KW gene therapy; chromosome 8.  
 XX  
 OS Homo sapiens.  
 XX  
 XX PN WO2003029271-A2.  
 XX PN  
 XX PD 10-APR-2003.  
 XX PD  
 XX PF 24-SEP-2002; 2002WO-US030474.  
 XX PF  
 XX PR 24-SEP-2001; 2001US-0324631P.  
 XX PR  
 XX PA (HYSE-) HYSEQ INC.  
 XX PA  
 XX PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
 XX PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
 XX PI Haley-Vicente D, Drmanac RT;  
 XX PI  
 XX PR WPI; 2003-371981/35.  
 XX PR  
 XX PR N-PSDB; ADC30829.  
 XX PR  
 XX PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
 XX PT treating conditions such as neurodegenerative diseases, anemias, platelet  
 XX PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 XX PT cancer.  
 XX  
 XX Claim 20; SEQ ID NO 1882; 1185pp; English.  
 XX  
 CC The invention relates to 971 novel human cDNA sequences (ADC29919-  
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
 CC invention also relates to nucleic acid sequences over 99% identical with  
 CC the novel human cDNAs. The invention additionally encompasses expression  
 CC vectors and host cells comprising a nucleic acid of the invention; the  
 CC recombinant production of a polypeptide of the invention; an antibody  
 CC against a polypeptide of the invention; a method of detecting  
 CC polynucleotides or polypeptides of the invention; and methods of  
 CC identifying a compound which binds to a polypeptide of the invention. The  
 CC invention further discloses methods of preventing, treating or  
 CC ameliorating a medical condition; kits comprising polynucleotide probes  
 CC and/or monoclonal antibodies for carrying out the methods of the  
 CC invention; methods for the identification of compounds that modulate the  
 CC expression or activity of the polynucleotide and/or polypeptide; and 767  
 CC contig sequences corresponding to the cDNA sequences of the invention  
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
 CC -ADC33394). The nucleic acids and polypeptides of the invention are  
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
 CC identification of mutations responsible for genetic disorders or other  
 CC traits, for assessing biodiversity, and in producing many other types of  
 CC data and products dependent on DNA and amino acid sequences. They are  
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, anaemia, platelet  
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 CC cancer. The nucleic acids may also be used as hybridisation probes or  
 CC primers, and in the recombinant production of a protein. The polypeptides

CC are also useful in generating antibodies, as molecular weight markers,  
 CC and as food supplements. The present sequence represents a specifically  
 CC claimed human polypeptide sequence of the invention. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 145 AA;  
 Query Match 84.1%; Score 645; DB 7; Length 145;  
 Best Local Similarity 83.4%; Pred. No. 1.1e-62;  
 Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGLHAGVLEDGP 60  
 DB 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGLHAGVLEDGP 60  
 QY 61 SSGVLRPAAGGIANPEKKNKCTGCPNQSLSGGLTKQKGLWTTAKDKAKMSAR 120  
 DB 61 PSNGVPRSTAGGIPNPEKTKNCETQCPNQSLSGGLTKQKGLWTTAKDKAKMPAK 120  
 QY 121 EVAISVTENIRQMDRSKRVTNCIN 145  
 DB 121 EVTINVTDSIQMDRSRRITKNCVN 145

RESULT 7  
 ADM46959  
 ID ADM46959 standard; protein; 145 AA.  
 XX AC ADM46959;  
 XX DT 03-JUN-2004 (first entry)  
 XX DE Brain and Acute Leukemia, Cytoplasmic alternate protein #1.  
 XX KW acute myelogenous leukemia; gene expression; BAALC;  
 XX KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 XX KW Cytoplasmic; exon.  
 XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 XX FT Misc-difference 41 /note= "encoded by GCS"  
 XX FT WO2003040347-A2.  
 XX PN 15-MAY-2003.  
 XX PD 12-NOV-2002; 2002WO-US036375.  
 XX PF 09-NOV-2001; 2001US-0348210P.  
 XX PR (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX PA Tanner SM, De La Chapell A;  
 XX PI WPI: 2003-441564/41.  
 XX DR N-PSDB; ADM46951.

XX Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 XX in a patient comprises assaying for the overexpression of one or more  
 XX BAALC transcripts in cells obtained from the patient.  
 XX PS Disclosure; SEQ ID NO 17; 78pp; English.

XX The invention relates to a method of characterizing acute myelogenous  
 XX leukemia (AML) in a patient by assaying for the overexpression of one or  
 XX more BAALC transcripts in cells obtained from the patient, where an  
 XX overexpression indicates that the patient has an aggressive form of AML.  
 XX The methods, kits and probes are useful for characterizing acute or  
 XX chronic myelogenous leukemia, or prostate cancer. They are also useful

CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.

XX SQ Sequence 145 AA;

Query Match 84.1%; Score 645; DB 7; Length 145;  
 Best Local Similarity 83.4%; Pred. No. 1.1e-62;  
 Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGLHAGVLEDGP 60  
 DB 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGLHAGVLEDGP 60  
 QY 61 SSGVLRPAAGGIANPEKKNKCTGCPNQSLSGGLTKQKGLWTTAKDKAKMSAR 120  
 DB 61 PSNGVPRSTAGGIPNPEKTKNCETQCPNQSLSGGLTKQKGLWTTAKDKAKMPAK 120  
 QY 121 EVAISVTENIRQMDRSKRVTNCIN 145  
 DB 121 EVTINVTDSIQMDRSRRITKNCVN 145

RESULT 8  
 ADO48475  
 ID ADO48475 standard; protein; 145 AA.

XX AC ADO48475;  
 XX DT 12-AUG-2004 (first entry)  
 XX DE Human PTH responsive gene protein.  
 XX KW PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
 XX KW transgenic animal; osteopathic; gene therapy; osteoporosis; human.

XX OS Homo sapiens.  
 XX PN WO2004044152-A2.  
 XX PD 27-MAY-2004.  
 XX PF 10-NOV-2003; 2003WO-US035655.  
 XX PR 12-NOV-2002; 2002US-0425532P.  
 XX PA (AMHP ) WYETH.  
 XX PI Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;  
 XX DR WPI: 2004-420299/39.  
 XX DR N-PSDB; ADO48474.

XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
 XX preparing a composition for diagnosing, treating or preventing bone  
 XX related disorders, e.g., osteoporosis.

XX Claim 9; SEQ ID NO 4; 169pp; English.

XX The invention relates to a novel PTH responsive gene (PAIGB) fragment  
 XX encoding a polypeptide. The invention further comprises: a chimeric  
 XX construct comprising the isolated nucleic acid fragment operatively  
 XX linked to suitable regulatory sequences; a host cell transformed with the  
 XX chimeric construct; a vector comprising the nucleic acid fragment;  
 XX obtaining a nucleic acid fragment encoding the polypeptide; a method for  
 XX obtaining a polypeptide; detecting the presence of the nucleic acid  
 XX fragment; an antibody that specifically binds to one or more epitopes of  
 XX a PAIGB polypeptide; a composition for regulating bone-forming activity  
 XX in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
 XX ; an agent that alters the expression of PAIGB gene or polypeptide;  
 XX determining whether an agent alters the expression of PAIGB mRNA;  
 XX screening agents for effectiveness in altering expression of the nucleic  
 XX acid fragment; screening for agents useful for treating bone related

CC disorders; evaluating the efficacy of a treatment of a bone related  
 CC disorder in a subject; identifying polypeptides capable of binding to  
 CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
 CC related agent; a transgenic animal comprising the DNA; an animal model  
 CC for the study of bone density modulation comprising a first group of  
 CC animals composed of the transgenic animal and a second group of control  
 CC animals; studying bone mass determinants; studying the modulation of bone  
 CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
 CC for treating bone related disorders; identifying whether an agent which  
 CC has bone forming activity; and a stably transfected cell line comprising  
 CC two constructs, the first construct comprising a ligand binding domain  
 CC linked to a DNA binding domain which is linked to an activation domain  
 CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopontin activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

XX Sequence 145 AA;

Query Match 84.1%; Score 645; DB 8; Length 145;  
 Best Local Similarity 83.4%; Pred. No. 1.1e-62;  
 Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MGCGRADAEIPRYESWTRETESTLWLTYSDALPSAAATSGPPEAGLHAGVLEDPG 60  
 DB 1 MGCGRADAEIPRYESWTRETESTLWLTYSDALPSAAATSGPPEAGLHAGVLEDPG 60  
 QY 61 SSGVLPAPAGGTANPEKKMNCCTQCPNSQSLSSGGLTKQNGLTWTEAKRDKRMSAR 120  
 DB 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPSQSLSSGGLTKQNGLTWTEAKRDKRMPAK 120  
 QY 121 EVAISVTENIRQMDRSKRVTNKCIN 145  
 DB 121 EVTINVTDSTQQMDRSRRITKNCVN 145

RESULT 9  
 ADM46961  
 ID ADM46961 standard; protein; 180 AA.

XX ADM46961;  
 XX 03-JUN-2004 (first entry)  
 XX Brain and Acute Leukemia, Cytoplasmic alternate protein #3.  
 XX acute myelogenous leukemia; gene expression; BAALC;  
 XX chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 XX Cytoplasmic; exon.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 XX Misc-difference 41 /note= "encoded by GCS"  
 XX WO2003040347-A2.  
 XX 15-MAY-2003.  
 XX 12-NOV-2002; 2002WO-US036375.  
 XX 09-NOV-2001; 2001US-0348210P.  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Tanner SM, De La Chapell A;  
 XX WPI; 2003-441564/41.  
 XX N-PSDB; ADM46955.

DR WPI; 2003-441564/41.  
 DR N-PSDB; ADM46953.  
 XX Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 PT in a patient comprises assaying for the overexpression of one or more  
 PT BAALC transcripts in cells obtained from the patient.  
 XX Disclosure; SEQ ID NO 19; 78pp; English.  
 XX The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
 CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.  
 XX Sequence 180 AA;

Query Match 80.5%; Score 617.5; DB 7; Length 180;  
 Best Local Similarity 67.2%; Pred. No. 1.6e-59;  
 Matches 121; Conservative 10; Mismatches 14; Indels 35; Gaps 1;  
 QY 1 MGCGRADAEIPRYESWTRETESTLWLTYSDALPSAAATSGPPEAGLHA----- 53  
 DB 1 MGCGRADAEIPRYESWTRETESTLWLTYSDALPSAAATSGPPEAGLHSLVLEAEKS 60  
 QY 54 -----GVLEDPGSPNGVLPAPAGGIANPEKKMNCCT 85  
 DB 61 KIKAPTDVSDGELFSASKMAPLAVFSGHMGLEDGLPSNGVPRSTAPGGIPNPEKTKNCET 120  
 QY 86 QCNSQSLSSGGLTKQNGLTWTEAKRDKRMSAREVAISVTENIRQMDRSKRVTNKCIN 145  
 DB 121 QCPNPSQSLSSGGLTKQNGLTWTEAKRDKRMPAKEVTINVTDSTQQMDRSRRITKNCVN 180

RESULT 10  
 ADM46963  
 ID ADM46963 standard; protein; 149 AA.

XX ADM46963;  
 XX 03-JUN-2004 (first entry)  
 XX Brain and Acute Leukemia, Cytoplasmic alternate protein #5.  
 XX acute myelogenous leukemia; gene expression; BAALC;  
 XX chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;  
 XX Cytoplasmic; exon.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 XX Misc-difference 41 /note= "encoded by GCS"  
 XX WO2003040347-A2.  
 XX 15-MAY-2003.  
 XX 12-NOV-2002; 2002WO-US036375.  
 XX 09-NOV-2001; 2001US-0348210P.  
 XX (OHIS ) UNIV OHIO STATE RES FOUND.  
 XX Tanner SM, De La Chapell A;  
 XX WPI; 2003-441564/41.  
 XX N-PSDB; ADM46955.

PT Characterizing acute or chronic myelogenous leukemia, or prostate cancer  
 PT in a patient comprising assaying for the overexpression of one or more  
 PT BAALC transcripts in cells obtained from the patient.

XX Disclosure; SEQ ID NO 21; 78pp; English.

XX The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
 CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.

XX Sequence 149 AA;

Query Match 61.7%; Score 473.5; DB 7; Length 149;  
 Best Local Similarity 66.0%; Pred. No. 9.8e-44;  
 Matches 95; Conservative 2; Mismatches 12; Indels 35; Gaps 1;

QY 1 MCGGSRADAIIPRYESWTRETESTWLTYSDDALPSAAATDGPAGGLHA----- 53  
 Db 1 MCGGSRADAIIPRYESWTRETESTWLTYSDDALPSAAATDGPAGGLHSLVLEAEKS 60  
 QY 54 -----GVLEDGSSNGVLRPAAPGGIANPEKMNCGT 85  
 Db 61 KIKAPTDSVSDGLFSASKMAPLAFVSHGWLDEGLPSNGVFRSTAPGGIPNPEKTKNCT 120  
 QY 86 QCPNSQSLSSGPLTKQKGLWTE 109  
 Db 121 QCPNPQSLSSGPLTKQKGLQTTE 144

RESULT 11

ADQ81902  
 ID ADQ81902 standard; protein; 92 AA.

AC ADQ81902;

DT 09-SEP-2004 (first entry)

XX Human dioxigenase 10.12.

XX Human; enzyme; dioxigenase 10.12; malignant tumour; inflammation;  
 KW immunological disease; haemopathy; HIV infection.

XX Homo sapiens.

XX CN1344798-A.

XX 17-APR-2002.

XX 29-SEP-2000; 2000CN-00125495.

XX 29-SEP-2000; 2000CN-00125495.

XX (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI; 2002-509506/55.

XX N-PSDB; ADQ81901.

XX New polypeptide human dioxigenase 10.12 and polynucleotides encoding this  
 PT polypeptide, useful for treating various diseases, such as malignant  
 PT tumors, inflammations, immunological diseases, hemopathy and HIV  
 PT infection.

PS Claim 1; SEQ ID NO 2; 33pp; Chinese.

XX The present invention discloses a new kind of polypeptide, human

CC dioxigenase 10.12, polynucleotides encoding this polypeptide, a DNA  
 CC recombination process to produce the polypeptide and antagonist against  
 CC the polypeptide. The present invention also discloses the method of  
 CC applying the polypeptide in treating various diseases, such as malignant  
 CC tumors, inflammations, immunological diseases, haemopathy and HIV  
 CC infection. The present sequence is the human dioxigenase 10.12.

XX Sequence 92 AA;

Query Match 43.7%; Score 335.5; DB 5; Length 92;  
 Best Local Similarity 72.7%; Pred. No. 8.4e-29;  
 Matches 64; Conservative 9; Mismatches 10; Indels 5; Gaps 1;

QY 58 DGPSSNGVLRPAAPGGIANPEKMNCGTQCPNSQSLSSGPLTKQKGLWTTAKRDAKRM 117  
 Db 10 DGPETGN-----APGGIPNPEKTKNCTQCPNPQSLSSGPLTKQKGLQTTEAKRDAKRM 64

QY 118 SAREVAISVTENIQMDRSKRVTKNCIN 145

Db 65 PAKEVTINVTDSIQMDRSRRITKNCVN 92

RESULT 12

ADQ48481

ID ADO48481 standard; protein; 54 AA.

AC ADO48481;

DT 12-AUG-2004 (first entry)

XX Mouse PTH responsive gene protein exon 2 splice variant.

XX PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
 KW transgenic animal; osteopathic; gene therapy; osteoporosis; mouse;  
 KW murine.

XX Mus sp.

XX WO2004044152-A2.

XX 27-MAY-2004.

XX 10-NOV-2003; 2003WO-US035655.

XX 12-NOV-2002; 2002US-0425532P.

XX (AMHP ) WYETH.

XX Robinson JA, Stojanovic-Susulic V, Babić P, Murrills RJ;

XX WPI; 2004-420299/39.

XX N-PSDB; ADO48480.

XX New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
 PT preparing a composition for diagnosing, treating or preventing bone  
 PT related disorders, e.g., osteoporosis.

XX Claim 9; SEQ ID NO 10; 169pp; English.

XX The invention relates to a novel PTH responsive gene (PAIGB) fragment  
 CC encoding a polypeptide. The invention further comprises: a chimeric  
 CC construct comprising the isolated nucleic acid fragment operatively  
 CC linked to suitable regulatory sequences; a host cell transformed with the  
 CC chimeric construct; a vector comprising the nucleic acid fragment;  
 CC obtaining a nucleic acid fragment encoding the polypeptide; a method for  
 CC obtaining a polypeptide; detecting the presence of the nucleic acid  
 CC fragment; an antibody that specifically binds to one or more epitopes of  
 CC a PAIGB polypeptide; a composition for regulating bone-forming activity  
 CC in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
 CC ; an agent that alters the expression of PAIGB gene or polypeptide;  
 CC determining whether an agent alters the expression of PAIGB mRNA;  
 CC screening agents for effectiveness in altering expression of the nucleic  
 CC acid fragment; screening for agents useful for treating bone related



CC disorders; evaluating the efficacy of a treatment of a bone related  
 CC disorder in a subject; identifying polypeptides capable of binding to  
 CC PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
 CC related agent; a transgenic animal comprising the DNA; an animal model  
 CC for the study of bone density modulation comprising a first group of  
 CC animals composed of the transgenic animal and a second group of control  
 CC animals; studying bone mass determinants; studying the modulation of bone  
 CC mass; studying an effect of PAIGB on bone disorders; identifying an agent  
 CC for treating bone related disorders; identifying whether an agent which  
 CC has bone forming activity; and a stably transfected cell line comprising  
 CC two constructs, the first construct comprising a ligand binding domain  
 CC linked to a DNA binding domain which is linked to an activation domain  
 CC all of which expression is driven by a constitutive promoter, the second  
 CC construct comprising multiple copies of DNA binding elements linked to a  
 CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
 CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
 CC polynucleotide has osteopontin activity. The PTH responsive gene may be  
 CC used to treat disorders by gene therapy. The nucleic acid is useful in  
 CC preparing a composition for diagnosing, treating or preventing bone  
 CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
 CC responsive gene protein of the invention.

XX SQ Sequence 54 AA;

Query Match 38.3%; Score 294; DB 8; Length 54;

Best Local Similarity 100.0%; Pred. No. 1.5e-24;

Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCGGSRADAIERYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAG 54

Db 1 MCGGSRADAIERYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAG 54

RESULT 13

ADM46962

ID ADM46962 standard; protein; 73 AA.

XX AC ADM46962;

XX DT 03-JUN-2004 (first entry)

XX DE Brain and Acute Leukemia, Cytoplasmic alternate protein #4.

XX KW acute myelogenous leukemia; gene expression; BAALC;

XX KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;

XX KW Cytoplasmic; exon.

XX OS Homo sapiens.

XX XX

XX Key Location/Qualifiers

XX FT Misc-difference 41

XX FT /note= "encoded by GCS"

XX WO2003040347-A2.

XX XX

XX PD 15-MAY-2003.

XX PF 12-NOV-2002; 2002WO-US036375.

XX XX

XX PR 09-NOV-2001; 2001US-0348210P.

XX XX

XX PA (OHIS ) UNIV OHIO STATE RES FOUND.

XX PI Tanner SM, De La Chapell A;

XX XX

XX DR WPI; 2003-441564/41.

XX DR N-PSDB; ADM46954.

XX XX

XX PT Characterizing acute or chronic myelogenous leukemia, or prostate cancer

XX PT in a patient comprises assaying for the overexpression of one or more

XX PT BAALC transcripts in cells obtained from the patient.

XX PS Disclosure; SEQ ID NO 20; 78pp; English.

XX The invention relates to a method of characterizing acute myelogenous  
 CC leukemia (AML) in a patient by assaying for the overexpression of one or  
 CC more BAALC transcripts in cells obtained from the patient, where an  
 CC overexpression indicates that the patient has an aggressive form of AML.  
 CC The methods, kits and probes are useful for characterizing acute or  
 CC chronic myelogenous leukemia, or prostate cancer. They are also useful  
 CC for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 CC (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 CC spliced RNA consisting of exons 1, 6 and 8.

XX SQ Sequence 73 AA;

Query Match 37.5%; Score 288; DB 7; Length 73;

Best Local Similarity 91.4%; Pred. No. 1.1e-23;

Matches 53; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MCGGSRADAIERYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLRD 58

Db 1 MCGGSRADAIERYESWTRETESTWLTYSDALPSAAATDSGPAGGLHAGVLRD 58

RESULT 14

ADM46960

ID ADM46960 standard; protein; 54 AA.

XX AC ADM46960;

XX DT 03-JUN-2004 (first entry)

XX DE Brain and Acute Leukemia, Cytoplasmic alternate protein #2.

XX KW acute myelogenous leukemia; gene expression; BAALC;

XX KW chronic myelogenous leukemia; prostate cancer; Brain and Acute Leukemia;

XX KW Cytoplasmic; exon.

XX OS Homo sapiens.

XX XX

XX Key Location/Qualifiers

XX FT Misc-difference 41

XX FT /note= "encoded by GCS"

XX WO2003040347-A2.

XX XX

XX PD 15-MAY-2003.

XX PF 12-NOV-2002; 2002WO-US036375.

XX XX

XX PR 09-NOV-2001; 2001US-0348210P.

XX PA (OHIS ) UNIV OHIO STATE RES FOUND.

XX PI Tanner SM, De La Chapell A;

XX XX

XX DR WPI; 2003-441564/41.

XX DR N-PSDB; ADM46952.

XX XX

XX PT Characterizing acute or chronic myelogenous leukemia, or prostate cancer

XX PT in a patient comprises assaying for the overexpression of one or more

XX PT BAALC transcripts in cells obtained from the patient.

XX PS Disclosure; SEQ ID NO 18; 78pp; English.

The invention relates to a method of characterizing acute myelogenous  
 leukemia (AML) in a patient by assaying for the overexpression of one or  
 more BAALC transcripts in cells obtained from the patient, where an  
 overexpression indicates that the patient has an aggressive form of AML.  
 The methods, kits and probes are useful for characterizing acute or  
 chronic myelogenous leukemia, or prostate cancer. They are also useful  
 for detecting BAALC overexpression. This sequence corresponds to a BAALC  
 (Brain and Acute Leukemia, Cytoplasmic) protein encoded by an alternative  
 spliced RNA consisting of exons 1, 6 and 8.

XX



SQ Sequence 54 AA;

Query Match 36.2%; Score 278; DB 7; Length 54;  
Best Local Similarity 94.4%; Pred. No. 8.8e-23;  
Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGGLHAG 54  
|||||  
Db 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGGLHSG 54  
|||||

RESULT 15  
ADO48477  
ID ADO48477 standard; protein; 54 AA.  
XX  
AC ADO48477;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE Human PTH responsive gene protein exon 2 splice variant.  
XX  
KW PTH responsive gene; PAIGB; bone-forming; bone; bone density modulation;  
KW transgenic animal; osteopathic; gene therapy; osteoporosis; human.  
XX  
OS Homo sapiens.  
XX  
FN WO200404152-A2.  
XX  
PD 27-MAY-2004.  
XX  
PF 10-NOV-2003; 2003WO-US035655.  
XX  
PR 12-NOV-2002; 2002US-0425532P.  
XX  
PA (AMHP ) WYETH.  
XX  
PI Robinson JA, Stojanovic-Susulic V, Babij P, Murrills RJ;  
XX  
WPI; 2004-420299/39.  
DR N-PSDB; ADO48476.  
XX  
New nucleic acid fragment encoding a PAIGB polypeptide, useful in  
preparing a composition for diagnosing, treating or preventing bone  
related disorders, e.g., osteoporosis.  
XX  
Claim 9; SEQ ID NO 6; 169pp; English.  
XX  
The invention relates to a novel PTH responsive gene (PAIGB) fragment  
encoding a polypeptide. The invention further comprises: a chimeric  
construct comprising the isolated nucleic acid fragment operatively  
linked to suitable regulatory sequences; a host cell transformed with the  
chimeric construct; a vector comprising the nucleic acid fragment;  
obtaining a nucleic acid fragment encoding the polypeptide; a method for  
obtaining a polypeptide; detecting the presence of the nucleic acid  
fragment; an antibody that specifically binds to one or more epitopes of  
a PAIGB polypeptide; a composition for regulating bone-forming activity  
in a mammal comprising the nucleic acid fragment, polypeptide or antibody  
; an agent that alters the expression of PAIGB gene or polypeptide;  
determining whether an agent alters the expression of PAIGB mRNA;  
screening agents for effectiveness in altering expression of the nucleic  
acid fragment; screening for agents useful for treating bone related  
disorders; evaluating the efficacy of a treatment of a bone related  
disorder in a subject; identifying polypeptides capable of binding to  
PAIGB; monitoring the effectiveness of treatment of a subject with a bone  
related agent; a transgenic animal comprising the DNA; an animal model  
for the study of bone density modulation comprising a first group of  
animals composed of the transgenic animal and a second group of control  
animals; studying bone mass determinants; studying the modulation of bone  
mass; studying an effect of PAIGB on bone disorders; identifying an agent  
for treating bone related disorders; identifying whether an agent which  
has bone forming activity; and a stably transfected cell line comprising  
two constructs, the first construct comprising a ligand binding domain  
linked to a DNA binding domain which is linked to an activation domain

CC all of which expression is driven by a constitutive promoter, the second  
CC construct comprising multiple copies of DNA binding elements linked to a  
CC minimal promoter which is linked to PAIGB cDNA, where upon the addition  
CC of chemical inducer, transcription of PAIGB gene is induced. The PAIGB  
CC polynucleotide has osteopathic activity. The PTH responsive gene may be  
CC used to treat disorders by gene therapy. The nucleic acid is useful in  
CC preparing a composition for diagnosing, treating or preventing bone  
CC related disorders, e.g., osteoporosis. This sequence represents a PTH  
CC responsive gene protein of the invention.  
XX  
SQ Sequence 54 AA;  
Query Match 36.2%; Score 278; DB 8; Length 54;  
Best Local Similarity 94.4%; Pred. No. 8.8e-23;  
Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGGLHAG 54  
|||||  
Db 1 MCGGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGGLHSG 54  
|||||

Search completed: November 17, 2004, 15:22:21  
Job time : 101 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 15:02:57 ; Search time 24.3333 Seconds  
(without alignments)  
395.183 Million cell updates/sec

Title: US-10-705-716A-2

Perfect score: 767

Sequence: 1 MCGGSGRAADIAIRYVESWT.....VTENIRQMDRSKRVTKNCIN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/ptodata/1/iaa/5A COMB.pap.\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B COMB.pap.\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A COMB.pap.\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B COMB.pap.\*
  - 5: /cgn2\_6/ptodata/1/iaa/PCUTS COMB.pap.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	92.5	12.1	778	6	Patent No. 5198347
2	84	11.0	726	4	Sequence 8115, Ap
3	83.5	10.9	700	3	Sequence 2, Appli
4	79.5	10.4	286	4	Sequence 176, Ap
5	79.5	10.4	305	4	Sequence 174, Ap
6	78.5	10.2	2516	3	Sequence 2, Appli
7	78.5	10.2	2516	3	Sequence 2, Appli
8	78.5	10.2	2516	4	Sequence 1, Appli
9	78	10.2	724	3	Sequence 10, Appli
10	78	10.2	885	4	Sequence 104, Appli
11	77.5	10.1	215	3	Sequence 473, Ap
12	77.5	10.1	445	4	Sequence 473, Ap
13	77.5	10.1	445	4	Sequence 473, Ap
14	77.5	10.1	445	4	Sequence 473, Ap
15	77.5	10.1	445	4	Sequence 473, Ap
16	77.5	10.1	445	4	Sequence 473, Ap
17	76.5	10.0	316	4	Sequence 16753, A
18	74.5	9.7	198	4	Sequence 23204, A
19	74.5	9.7	650	4	Sequence 469, Ap
20	74.5	9.7	650	4	Sequence 469, Ap
21	74.5	9.7	650	4	Sequence 469, Ap
22	74.5	9.7	650	4	Sequence 469, Ap
23	74.5	9.7	650	4	Sequence 469, Ap
24	74.5	9.7	743	4	Sequence 494, Ap
25	74.5	9.7	743	4	Sequence 494, Ap
26	74.5	9.7	1002	4	Sequence 475, Ap
27	74.5	9.7	1002	4	Sequence 475, Ap

74.5 9.7 1002 4 US-09-834-759-475 Sequence 475, App  
74.5 9.7 1002 4 US-09-590-751A-475 Sequence 475, App  
74.5 9.7 1095 4 US-09-620-405B-493 Sequence 493, App  
74.5 9.7 1095 4 US-09-834-759-493 Sequence 493, App  
74 9.6 244 4 US-09-252-991A-18465 Sequence 18465, A  
74 9.6 715 4 US-09-252-991A-32740 Sequence 32740, A  
73.5 9.6 172 4 US-09-252-991A-17765 Sequence 17765, A  
73.5 9.6 512 4 US-09-451-739H-16 Sequence 16, Appli  
73.5 9.6 3623 4 US-09-341-461-2 Sequence 2, Appli  
73 9.5 774 4 US-09-252-991A-16789 Sequence 16789, A  
72.5 9.5 139 4 US-09-252-991A-19445 Sequence 19445, A  
72.5 9.5 435 4 US-09-270-767-43092 Sequence 43092, A  
72.5 9.5 583 4 US-09-489-039A-7653 Sequence 7653, Ap  
72.5 9.5 1184 4 US-10-140-002-394 Sequence 394, App  
72 9.4 217 4 US-09-252-991A-26962 Sequence 26962, A  
71.5 9.3 240 4 US-09-252-991A-21991 Sequence 21991, A  
71.5 9.3 266 4 US-09-252-991A-27673 Sequence 27673, A  
71.5 9.3 675 1 US-08-386-495-10 Sequence 10, Appli

## ALIGNMENTS

RESULT 1  
5198347-4  
; Patent No. 5198347  
; APPLICANT: MILLER, LOUIS H.; ADAMS, JOHN H.; KASLOW,  
; DAVID C.; FANG, XIANGDOUG  
; TITLE OF INVENTION: DNA ENCODING PLASMIDIUM VIVAX AND  
; PLASMIDIUM KNOWLES DUFFY RECEPTOR  
; NUMBER OF SEQUENCES: 27  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/554,837  
; FILING DATE: 20-JUL-1990  
; SEQ ID NO:4:  
; LENGTH: 778  
5198347-4

Query Match 12.1%; Score 92.5; DB 6; Length 778;  
Best Local Similarity 33.0%; Pred. No. 0.11; Indels 3; Gaps 3;  
Matches 29; Conservative 12; Mismatches 12

Qy 31 TSDALPSAAATDSGPAGGLHAGVLPSSNGV-LRPAAPGGIANPEKKNMC-GTQCP 88  
Db 352 TVSSDVSFVGKDSGSPSTASASHALAGENGVEVHNTDTPEKDEGKADPQKDIEVKGQDT 411  
Qy 89 NSQLSS-GPLTKQKGLWTEAKRDAK 115  
Db 412 DRSQSGSLGPHTDERTATLGETHMEKDT 439

RESULT 2  
US-09-328-352-8115  
; Sequence 8115, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: GARY L. BRETON ET AL.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 8115  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-8115

Query Match 11.0%; Score 84; DB 4; Length 726;  
Best Local Similarity 35.1%; Pred. No. 0.96;  
Matches 27; Conservative 8; Mismatches 22; Indels 20; Gaps 4;

QY 2 GCGSRADAEIPRYESWTRETESTWLTVDSDALPSA--AAYDSG-----PRAGGLHAG 54  
Db 181 GFGAGREDVWPNDVNWGDKE--MLAHNSEALAGSNLAATEMGLIYVNP-----231  
QY 55 VLEDGSSNGVLRPAAP 71  
Db 232 ----GQASGDPRSAP 244

## RESULT 3

US-09-408-647A-2  
; Sequence 2, Application US/09408647A  
; Patent No. 6399858  
; GENERAL INFORMATION:  
; APPLICANT: Kobayashi, Donald  
; TITLE OF INVENTION: Chitinase Gene from Stenotrophomonas  
; FILE REFERENCE: Rut-Cook 98-0090  
; CURRENT APPLICATION NUMBER: US/09/408,647A  
; PRIOR FILING DATE: 1999-08-26  
; PRIOR APPLICATION NUMBER: 60/098,036  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 700  
; TYPE: PRT  
; ORGANISM: Stenotrophomonas maltophilia  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(41)  
; NAME/KEY: DOMAIN  
; LOCATION: (196)...(290)  
; NAME/KEY: DOMAIN  
; LOCATION: (330)...(483)  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: Genbank No. 6399858 AF014950  
; DATABASE ENTRY DATE: 1997-09-23  
US-09-408-647A-2

Query Match 10.9%; Score 83.5; DB 3; Length 700;  
Best Local Similarity 24.0%; Pred. No. 1;  
Matches 31; Conservative 14; Mismatches 57; Indels 27; Gaps 2;  
QY 26 TWLTYTDSALPSAATDSGPEAGGLHAGVLEDGSPSSNGVLRPAAPGGIANPEKKM----81  
Db 159 TWANASAGSHTPKAVATDNNATVSSATVSVTVTASSNDTTPSPVGGGLASPSKTATTVN 218  
QY 82 -----NCG-----TQCPNSQSLSSGPIQTQKQNGLWTTTEAKRDKRWS 118  
Db 219 LVMSAATDNSGGSGVAGDYVRNGSLVGSFSAQTQYTDGGLTASTAYTYTVRARDNAGNAS 278  
QY 119 AREVAISVT 127  
Db 279 AQSGSISVT 287

## RESULT 4

US-09-071-035-176  
; Sequence 176, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 176:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 286 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-176

Query Match 10.4%; Score 79.5; DB 4; Length 286;  
Best Local Similarity 26.6%; Pred. No. 0.83; Indels 41; Gaps 9;  
Matches 47; Conservative 16; Mismatches 73; Indels 41; Gaps 9;  
QY 3 CGG-----GRADAIEPRYVESW-----TRETESTWLTVDSDALPSAAAT-DSGPEAG 49  
Db 1 CGGKSTENTDSRSSAAESTTVESTKASATKSSK-ATTKSSDAKPSGTTTADSKATAS 59  
QY 50 GLHAGVLEDGSPSSNGVLRPAAPGGIANPEKKM-COTQCPN---SQSLSSGPIQTQKQNGL 105  
Db 60 STKEA-----ANNGSAEKQSPAKNANPDQANQVNLNLANMPPGQGLPQAILTSQTNF 113  
QY 106 WTEAKRDKRMSAREVAISVTENTRQMD-RSKRVT-----KNCIN 145  
Db 114 LTAATTSQADQNFRVLYAEKAIQVNDARVNQLTPISSEFKKTYGSDAEAKNAN 170

## RESULT 5

US-09-071-035-174  
; Sequence 174, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes

```

; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 174:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-174

```

```

Query Match 10.4%; Score 79.5; DB 4; Length 305;
Best Local Similarity 26.6%; Pred. No. 0.91;
Matches 47; Conservative 16; Mismatches 73; Indels 41; Gaps 9;

```

```

QY 3 CGG-----SRDAIEPRVYESW-----TRETSTLTWTTSDALPSAAAT-DGSGPEAG 49
DB 20 CGGKSTNTDSRSSAESITVESTKASATKSSK-ATTKSSDAKPSGTTTADSKATAS 78
QY 50 GLHAGVLEDGSPSNGVLRPAAPGGIANPEKKN-CGTQCPN---SQSLSSGPLETKQKGL 105
DB 79 STKEA-----ANNGSAEKQSPAKNANPDQANVLNQLANNFPQGLPQAILTSQTNNF 132
QY 106 WTTEAKRDKMSAREVAISVTENIRQMD-RSKRVT-----KNCIN 145
DB 133 LTAATTSQDQNNFRVLYAEKBAIPVNDARVNQLTPISSFEKTKYGSDAEAKNAVN 189

```

```

RESULT 6
US-08-374-077C-2
; Sequence 2, Application US/08374077C
; Patent No. 6027912
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Invertebrate Algal
; TITLE OF INVENTION: Calcium Channel Subunit
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,077C
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-374-077C-2

```

```

Query Match 10.2%; Score 78.5; DB 3; Length 2516;
Best Local Similarity 22.4%; Pred. No. 25;
Matches 39; Conservative 21; Mismatches 61; Indels 53; Gaps 8;

```

```

QY 3 CGGSRADAIEPRY--YESWTRETSTLTWTTSDALPSAAATDSGPEAGLHAGVLEDGP 60
DB 207 CGGGISAPPPRLTPEEAWQLQPQ-----NSVTSAGSTNSFSFGG-----GR 249
QY 61 SSGVLRPAAPGGIANPEKKNK-----GTQC-----PNS 90
DB 250 DNSSY--SAVGDSSSSSNSCNDITGUNTSLHGLGVGDVCSFIADCDNDEDDDDGDDNN 307
QY 91 QSLSSGPLE-TQKQGLWTTEAKRDKMSAREVAISVTENIRQMDRSKRVTKNC 143
DB 308 QDLSSTQLRTAAIIVAAVAAAQKQAQCSLADCE-SFSDRQDADQDVRRIQDC 360

```

```

RESULT 7
US-08-895-590-2
; Sequence 2, Application US/088955590
; Patent No. 6207410
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,590
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,888
; FILING DATE: 19-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-263
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-895-590-2

```

```

Query Match 10.2%; Score 78.5; DB 3; Length 2516;
Best Local Similarity 22.4%; Pred. No. 25;
Matches 39; Conservative 21; Mismatches 61; Indels 53; Gaps 8;

```

```

QY 3 CGGSRADAIEPRY--YESWTRETSTLTWTTSDALPSAAATDSGPEAGLHAGVLEDGP 60
DB 207 CGGGISAPPPRLTPEEAWQLQPQ-----NSVTSAGSTNSFSFGG-----GR 249
QY 61 SSGVLRPAAPGGIANPEKKNK-----GTQC-----PNS 90

```

```

307
250 DNSSY--SAVGDDSSSSNCNDITGDNSTLHGLGVGDCVSFIADCDNSDDDDGDPNN
QY 91 QSLSSGPL-TQKQGLWTTAKRDAKRMSAREVAISVTENIQMDRSKRVTKNC 143
Db 308 QDLSSQTLRTAAIVAAVAANAQAQEQSLADCE-SFSDRRQDAEDVRIIQDC 360

RESULT 8
US-09-539-879A-2
; Sequence 2, Application US/09539879A
; Patent No. 6436627
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; Ren, Dejian
; Zheng, Wei
; Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha
; Calcium Channel Subunit
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/539,879A
; FILING DATE: 31-Mar-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/111,865
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/374,077
; FILING DATE: 19-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-539-879A-2

Query Match 10.2%; Score 78.5; DB 4; Length 2516;
Best Local Similarity 22.4%; Pred. No. 25;
Matches 39; Conservative 21; Mismatches 61; Indels 53; Gaps 8

QY 3 CGGSRAAIEPRY--YESWTRETSTWLTYTDSALPSAATDSGPAGGLHAGVLEDP 60
Db 207 CGGGISAPPRLTPEAWQLQPQ-----NSVTSAGSTNSFSGG-----GR 249

QY 61 SSVGVLRAAPGGTANPEKKNK-----GTQC-----PNS 90
Db 250 DNSSY--SAVGDDSSSSNCNDITGDNSTLHGLGVGDCVSFIADCDNSDDDDGDPNN 307

QY 91 QSLSSGPL-TQKQGLWTTAKRDAKRMSAREVAISVTENIQMDRSKRVTKNC 143
Db 308 QDLSSQTLRTAAIVAAVAANAQAQEQSLADCE-SFSDRRQDAEDVRIIQDC 360

```

```
/ CURRENT APPLICATION NUMBER: US/09/220,528A
/ CURRENT FILING DATE: 1998-12-24
/ EARLIER APPLICATION NUMBER: 09/218,698
/ EARLIER FILING DATE: 1998-12-22
/ EARLIER APPLICATION NUMBER: 60/108,148
/ EARLIER FILING DATE: 1998-11-12
/ EARLIER APPLICATION NUMBER: 09/163,283
/ EARLIER FILING DATE: 1998-09-29
/ NUMBER OF SEQ ID NOS: 120
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 104
/ LENGTH: 215
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-220-528-104

Query Match          10.1%; Score 77.5; DB 3; Length 215;
Best Local Similarity 26.5%; Pred. No. 0.93;
Matches 31; Conservative 12; Mismatches 41; Indels 33; Gaps 5;

QY 5 GSRDAIEPR-----YYESWTRETESTWLTYSDALPSAAATDSGPEAGGLHA----- 53
DB 65 GARAALGQGEAGAACRSCARSAWATPTWCWSASAAAPAAAPAA--LHTTSWAP 122

QY 54 GVLEDGSSNGVLRPAAPGGIANPEKMMCGTQCPNSQSLSGPLTKQKONGLWTTTEA 110
DB 123 AYWAPGPCD---RPRAPGPSASP-----AADPRATKRPSPWTTSTA 159

RESULT 12
US-09-620-405B-473
/ Sequence 473, Application US/09620405B
/ Patent No. 6528054
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Yuqiu
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Hepler, William T.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.470C8
/ CURRENT APPLICATION NUMBER: US/09/620,405B
/ CURRENT FILING DATE: 2000-07-20
/ NUMBER OF SEQ ID NOS: 495
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 473
/ LENGTH: 445
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-620-405B-473

Query Match          10.1%; Score 77.5; DB 4; Length 445;
Best Local Similarity 21.5%; Pred. No. 2.7;
Matches 37; Conservative 28; Mismatches 58; Indels 49; Gaps 7;

QY 7 RADAIEP-----RYVESWTRETESTWLTYSDDLPSAAATDSGPEAGGLHAGVLEDGP 60
DB 78 RADELPLPSESQKQDYESSWSESICETVSKQDVCLPKAAHQKEIDKIN-----GKLEESP 133

QY 61 SSVGVLPRPAAPGGIANPEKMM-----NCGTQCPNS 90
DB 134 DNDGFLKAPCRMVSIPTKALELMDMTFKAEPPKPSAFPAIEQMKSVPNKALELKNE 193

QY 91 QSLSSG---PLTKQKONGL-----WTEAKRDAKMSAREVAISVTENIQMDR 135
DB 194 QTLRADQMFPPSSKQKVEENSWSDESILRET--VSQKDVCPVCPKATHQKEMDK 243

RESULT 13
US-09-433-826B-473
/ Sequence 473, Application US/09433826B
/ Patent No. 6528054
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Yuqiu
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Hepler, William T.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.470C8
/ CURRENT APPLICATION NUMBER: US/09/620,405B
/ CURRENT FILING DATE: 2000-07-20
/ NUMBER OF SEQ ID NOS: 495
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 473
/ LENGTH: 445
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-433-826B-473

Query Match          10.1%; Score 77.5; DB 4; Length 445;
Best Local Similarity 21.5%; Pred. No. 2.7;
Matches 37; Conservative 28; Mismatches 58; Indels 49; Gaps 7;

QY 7 RADAIEP-----RYVESWTRETESTWLTYSDDLPSAAATDSGPEAGGLHAGVLEDGP 60
DB 78 RADELPLPSESQKQDYESSWSESICETVSKQDVCLPKAAHQKEIDKIN-----GKLEESP 133

QY 61 SSVGVLPRPAAPGGIANPEKMM-----NCGTQCPNS 90
DB 134 DNDGFLKAPCRMVSIPTKALELMDMTFKAEPPKPSAFPAIEQMKSVPNKALELKNE 193
```

```
/ Patent No. 6579973
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Yuqiu
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Harlocker, Susan L.
/ TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
/ FILE REFERENCE: 210121.470C4
/ CURRENT APPLICATION NUMBER: US/09/433,826B
/ CURRENT FILING DATE: 1999-11-03
/ NUMBER OF SEQ ID NOS: 474
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 473
/ LENGTH: 445
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-433-826B-473

Query Match          10.1%; Score 77.5; DB 4; Length 445;
Best Local Similarity 21.5%; Pred. No. 2.7;
Matches 37; Conservative 28; Mismatches 58; Indels 49; Gaps 7;

QY 7 RADAIEP-----RYVESWTRETESTWLTYSDDLPSAAATDSGPEAGGLHAGVLEDGP 60
DB 78 RADELPLPSESQKQDYESSWSESICETVSKQDVCLPKAAHQKEIDKIN-----GKLEESP 133

QY 61 SSVGVLPRPAAPGGIANPEKMM-----NCGTQCPNS 90
DB 134 DNDGFLKAPCRMVSIPTKALELMDMTFKAEPPKPSAFPAIEQMKSVPNKALELKNE 193

RESULT 14
US-09-604-287A-473
/ Sequence 473, Application US/09604287A
/ Patent No. 6586572
/ GENERAL INFORMATION:
/ APPLICANT: Jiang, Yuqiu
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Hepler, William T.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.470C7
/ CURRENT APPLICATION NUMBER: US/09/604,287A
/ CURRENT FILING DATE: 2000-06-22
/ NUMBER OF SEQ ID NOS: 489
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 473
/ LENGTH: 445
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-604-287A-473

Query Match          10.1%; Score 77.5; DB 4; Length 445;
Best Local Similarity 21.5%; Pred. No. 2.7;
Matches 37; Conservative 28; Mismatches 58; Indels 49; Gaps 7;

QY 7 RADAIEP-----RYVESWTRETESTWLTYSDDLPSAAATDSGPEAGGLHAGVLEDGP 60
DB 78 RADELPLPSESQKQDYESSWSESICETVSKQDVCLPKAAHQKEIDKIN-----GKLEESP 133

QY 61 SSVGVLPRPAAPGGIANPEKMM-----NCGTQCPNS 90
DB 134 DNDGFLKAPCRMVSIPTKALELMDMTFKAEPPKPSAFPAIEQMKSVPNKALELKNE 193
```

Search completed: November 17, 2004, 15:30:02  
Job time : 25.3333 secs







```
US-10-293-239-18
; Sequence 18, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-18
Query Match      36.2%; Score 278; DB 14; Length 54;
Best Local Similarity 94.4%; Pred. No. 2.1e-20;
Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MCGGSRADAIEPRYESWTRETESTWLTYTDSALPSSAAATDSGPEAGGLHAG 54
Db 1 MCGGSRADAIEPRYESWTRETESTWLTYTDSALPSSAAAPDSGPEAGGLHSG 54

RESULT 7
US-10-293-239-22
; Sequence 22, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-22
Query Match      35.5%; Score 272; DB 14; Length 80;
Best Local Similarity 94.3%; Pred. No. 1.4e-19;
Matches 50; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MCGGSRADAIEPRYESWTRETESTWLTYTDSALPSSAAATDSGPEAGGLHA 53
Db 1 MCGGSRADAIEPRYESWTRETESTWLTYTDSALPSSAAAPDSGPEAGGLHS 53

RESULT 8
US-10-293-239-37
; Sequence 37, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-37
Query Match      18.3%; Score 140; DB 14; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.2e-07;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 RADAIEPRYESWTRETESTWLTYT 31
Db 1 RADAIEPRYESWTRETESTWLTYT 25

RESULT 9
US-10-293-239-35
; Sequence 35, Application US/10293239
; Publication No. US20030119043A1
; GENERAL INFORMATION:
; APPLICANT: Tanner, Stephan
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: BAALC expression as a diagnostic marker for acute leukemia
; FILE REFERENCE: 22727/04101
; CURRENT APPLICATION NUMBER: US/10/293,239
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/348,210
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-239-35
Query Match      12.9%; Score 99; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.0055;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DAIEPRYESWTRETEST 26
Db 1 DAIEPRYESWTRETEST 18

RESULT 10
US-10-767-701-48905
; Sequence 48905, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 48905
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3476-020-P1-K1-A12.pep
US-10-767-701-48905
Query Match      12.3%; Score 94; DB 16; Length 147;
Best Local Similarity 27.5%; Pred. No. 0.24;
Matches 30; Conservative 17; Mismatches 52; Indels 10; Gaps 3;

Qy 36 LPSSAAATDSGPEAGGLHAGVLEDFSSNGVLRPAAPGIANPEKKMNCGTCCPNSQS--- 92
```

## RESULT 11

37 PSAAATDSGPEAGGLHAGVLEDGP-SSNGVLRPAAPGGIANPEKKMNCGTQCPN---SQS 92

Qv	93	LSSGPLTQKONG	-----LWTE-----AKRDAKRMSA-----	119
----	----	--------------	-------------------------------	-----

Qy	32	DSDALPSAAATD	GPEAGGLHAGVLDG	PSSNGVLRAAPGGI	--ANPEKMNCGTQP	88
Db	147	DSDDAFAAAAHD	DDPAAVWAAGGLGGEN	SSSG-LPPAAGAATAAE	EPSPLSLGLPLP	205

Qy 89 -----NSQSLSGELTQKQGLWTTTEAKRDAKMSAREV 122

206 AAEPAEAAADESRNSQQAS--P LLEGE G--NAQLLAWVRMVR EEV 250

RESULT 14

```

RESULT 14
US-10-437
; Sequenc
; Publica
; GENERAL
; APPLIC
; APPLIC
; APPLIC
; APPLIC
; APPLIC
; APPLIC
; APPLIC
; APPLIC
; TITLE
; TITLE
; FILE R
; CURREN
; CURREN
; NUMBE
; SEQ ID

```

```

; NAME: TAX
; ORGANISM:
; FEATURE:

```

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_38593C.1.pap  
US-10-437-963-137075

Query Match 11.1%; Score 85; DB 16; Length 450;  
Best Local Similarity 25.6%; Pred. No. 8;  
Matches 45; Conservative 14; Mismatches 53; Indels 64; Gaps 8;

Qy 4 GGSRADAI-----EPYYESWTRETESTWLTYTDSALPSAAATDSGPEAGGLH 52  
Db 3 GGS-ADATKEMEALLVQNFNAVSGTCETSSKGGKVDNSGSHSPEDDDDAQG-- 59  
Qy 53 AGVLDDGSSNGVLPAAPGGIANPEKMNCGTCQPNQSLSGSLTQ----- 100  
Db 60 -----DGPSQDG-----GSEAKKKKKK-----SKSKKKKGLQQTDPSPIDELF 101  
Qy 101 -----KONGLM--TTEAKRDAKM-----SAREVAISVTENIRQMDRS 136  
Db 102 PSGDFPEGEIOQYKDDNLWRTTSEEKRELERLQKPMYNAVRAAEVHRQVRKHMRS 157

RESULT 15

US-10-282-122A-45268  
; Sequence 45268, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA 034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 45268  
; LENGTH: 718  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-10-282-122A-45268

Query Match 11.0%; Score 84; DB 15; Length 718;  
Best Local Similarity 35.1%; Pred. No. 18;  
Matches 27; Conservative 8; Mismatches 22; Indels 20; Gaps 4;

Qy 2 GCGSRADAIEPRYYESWTRETESTWLTYTDSALPSA--AATDSG-----PEAGGLHAG 54  
Db 173 GFGAGREDVWEPDNDVNWGDEKE--WLAHRNSEALAGSNLAATENGLIYVNP----- 223  
Qy 55 VLEDGPPSSNGVLRPAAAP 71  
Db 224 -----GPOASGDPSPSAAP 236

Search completed: November 17, 2004, 15:48:52  
Job time : 93.3333 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 15:00:21 ; Search time 19.6667 Seconds  
(without alignments)  
709.395 Million cell updates/sec

Title: US-10-705-716a-2  
Perfect score: 767  
Sequence: 1 MCGGSRADAIEPRYYESWT.....VTENIQMDRSKRVTKNKCN 145

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 9621673 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92.5	12.1	571	2 T43456	hypothetical prote
2	92.5	12.1	778	2 A35970	erythrocyte-bindin
3	88	11.5	365	1 GNVSSC	genome polyprotein
4	83	10.8	1433	2 S54587	CAT8 protein - yea
5	81	10.6	269	2 T37073	hypothetical prote
6	81	10.6	967	2 S66852	hypothetical prote
7	79.5	10.4	2559	2 T09144	probable guanine n
8	78.5	10.2	1199	2 A40670	nuclear envelope p
9	78	10.2	885	2 T09225	A kinase anchor pr
10	78	10.2	3488	2 T34418	hypothetical prote
11	76.5	10.0	600	2 C69899	conserved hypotet
12	76	9.9	645	2 S19156	serotonin receptor
13	76	9.8	839	2 B84824	hypothetical prote
14	75.5	9.8	499	2 S22571	hypothetical prote
15	75.5	9.8	1063	2 T03743	integrase-like pro
16	75.5	9.8	1122	2 T47424	bifocal protein -
17	75	9.8	368	1 TVMSML	hypothetical prote
18	75	9.8	521	2 I51693	transforming prote
19	74.5	9.7	601	2 AH0784	XPolycomb - Africa
20	74.5	9.7	4957	2 T03455	probable transport
21	74.5	9.7	5262	2 T03454	ALR protein - huma
22	74	9.6	260	2 B38594	ALR protein - huma
23	74	9.6	742	2 T38001	tropoin I - fruit
24	74	9.6	832	2 T31878	probable phosphati
25	74	9.6	962	2 T00262	hypothetical prote
26	73.5	9.6	876	1 A57988	hypothetical prote
27	73.5	9.6	960	2 T37916	regulatory protein
28	73.5	9.6	1575	2 S68448	probable heterochr
29	73	9.5	403	2 H98327	synaptojanin, 170K
30					enantiomer-selecti

30	73	9.5	453	2 AE2955	glutamyl-tRNA amid
31	73	9.5	781	2 T41551	hypothetical prote
32	73	9.5	1274	2 T37193	enamelin matrix pr
33	72.5	9.5	251	2 C75521	cytochrome-related
34	72.5	9.5	284	2 I51172	transcription fact
35	72.5	9.5	435	2 AG1028	prepilin (imported
36	72.5	9.5	503	2 T35053	probable solute-bi
37	72.5	9.5	516	2 JE0301	inulinase (EC 3.2.
38	72.5	9.5	1343	2 AF0611	cell division prot
39	72.5	9.5	3623	2 T08618	microtubule-associ
40	72.5	9.5	5327	2 T13564	intrinsic factor-B
41	72	9.4	108	2 T31565	hypothetical prote
42	72	9.4	1038	1 JC5757	DNA-directed DNA p
43	72	9.4	1038	2 T18222	DNA polymerase del
44	72	9.4	1573	2 T50113	3-dehydroquinase s
45	71.5	9.3	381	2 AB3048	8-amino-7-oxononan

## ALIGNMENTS

## RESULT 1

T43456  
hypothetical protein DKFp434L061.1 - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T43456  
Ripoustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z22516  
A:Accession: T43456  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-571 <AAA>  
A:Cross-references: UNIPROT:O75175; EMBL:AL133647  
A:Experimental source: adult testis; clone DKFZp434L061  
C:Genetics:  
A:Note: DKFZp434L061.1  
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 12.1%; Score 92.5; DB 2; Length 571;  
Best Local Similarity 31.9%; Pred. No. 0.95;  
Matches 30; Conservative 7; Mismatches 32; Indels 25; Gaps 3;

QY	31	TDSDALPSAAATDSPEAGGLH-----AGVLEDGPSNGVLRAAP--	71
DB	111	TDSEVSQSPAKNGSKPVHSNQHPQSPVPTYPGPPPAASALSTTPGNGVPAPAPPS	170
QY	72	--GGIANPEKMKNCGTQCPNSQSLs-----SGPLT	99
DB	171	ALGPKASPSPSHSGTPAPYAQAQAVAPAPSPGST	204

## RESULT 2

A35970  
erythrocyte-binding protein - Plasmodium knowlesi  
C:Species: Plasmodium knowlesi  
C:Date: 14-Dec-1990 #sequence\_revision 14-Dec-1990 #text\_change 09-Jul-2004  
C:Accession: A35970  
R:Adams, J.H.; Hudson, D.E.; Torii, M.; Ward, G.E.; Wellems, T.E.; Alkawa, M.; Miller, L.  
Cell 63, 141-153, 1990  
A:Title: The Duffy receptor family of plasmodium knowlesi is located within the micronem  
A:Reference number: A35970; MUID:91004213; PMID:2170017  
A:Accession: A35970  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-778 <ADA>  
A:Cross-references: UNIPROT:P22545; GB:M68518; GB:M37513; NID:g160273; PID:g160274  
C:Keywords: transmembrane protein

Query Match 12.1%; Score 92.5; DB 2; Length 778;  
Best Local Similarity 33.0%; Pred. No. 1.4;  
Matches 29; Conservative 12; Mismatches 44; Indels 3; Gaps 3;

QY 31 TSDALPSAAATDSGPEAGLHAGVLEDPSSNGV-LRPAAPGGIANPEKKNVC-GTQCP 88  
Db 352 TVSSDVPVSGKDSGSTSASHALAGEVHVHTDTEPKEDGKADPQDIEVKGQDT 411

QY 89 NSQSLSS-GPLTKQKGLWTTEAKRDAK 115  
Db 412 DRSQSLGHTDTRATLGHTEKOTE 439

RESULT 3  
GNVSSC  
genome polyprotein - sugarcane mosaic virus (strain SC) (fragment)  
N;Contains: carboxyl end of nuclear inclusion protein b; coat protein  
C;Species: sugarcane mosaic virus, SCMV  
C;Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 09-Jul-2004  
C;Accession: PH0207  
R;Frenkel, M.J.; Jilka, J.M.; McKern, N.M.; Strike, P.M.; Clark Jr., J.M.; Shukla, D.D.;  
J. Gen. Virol. 72, 237-242, 1991  
A;Title: Unexpected sequence diversity in the amino-terminal ends of the coat proteins of  
A;Reference number: PH0207; MUID:91132116; PMID:1993866  
A;Accession: PH0207  
A;Molecule type: genomic RNA  
A;Residues: 1-365 <PRE>  
A;Cross-references: UNIPROT:P25242; GB:D00948; NID:g222123; PIDN:BAA00796.1; PID:g222124  
C;Superfamily: tobacco etch virus genome polyprotein  
C;Keywords: coat protein; inclusion protein  
F;1-52/Product: nuclear inclusion protein b (fragment) #status predicted <IPB>  
F;53-365/Product: coat protein #status predicted <COA>

Query Match 11.5%; Score 88; DB 1; Length 365;  
Best Local Similarity 21.6%; Pred. No. 1.5;  
Matches 30; Conservative 24; Mismatches 79; Indels 6; Gaps 4;

QY 5 GSRADATEPRYYSWTRSTWLTYSDALPSAAATDSGPEAGLHAGVLEDPSS 62  
Db 21 GIKKEEIE-KYFKQFAKDLPGYLEYND-EVFHQAGTVDAAGQGGGNAGTQPATGAAA 78

QY 63 NGVLRPAAPGGIANPEKKNVC-GTQCPNSQSLSSGPLTKQKGLWTTEAKRDAKMSAREV 122  
Db 79 QGGAQPPATGAAGPPTTQ--GSQLPQGGATGGGAGTGGTGSVTTGGORDKVDAGTT 136

QY 123 AISVTENIRQWDRSKRVTK 141  
Db 137 GKITVPKLAAMKQWLPK 155

RESULT 4  
S54587  
CAT8 protein - yeast (Saccharomyces cerevisiae)  
N;Alternate names: MSP8 protein; protein YM8021.06c; protein YMR280C  
C;Species: Saccharomyces cerevisiae  
C;Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 16-Aug-2004  
C;Accession: S54587; S48234; S61595; S49498  
R;Pearson, D.; Bowman, S.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: S54587  
A;Accession: S54587  
A;Molecule type: DNA  
A;Residues: 1-1433 <PEA>  
A;Cross-references: UNIPROT:P39113; EMBL:Z49704; NID:g825540; PIDN:CAA89778.1; PID:g8255  
A;Experimental source: strain AB972  
R;Grzesitza, D.  
submitted to the EMBL Data Library, March 1994  
A;Reference number: S48234  
A;Accession: S48234  
A;Molecule type: DNA  
A;Residues: 1-746, 'L', 748-1433 <GRZ>  
A;Cross-references: EMBL:X78344; NID:g559523; PIDN:CAA55139.1; PID:g559524  
R;Boles, E.; Hettmann, C.; Zimmermann, F.K.  
submitted to the EMBL Data Library, December 1995  
A;Reference number: S61594  
A;Accession: S61595

QY 11 IEPRYVESWT-----RETSTWLTYSDALPSAAATDSGPEAGLHAGVLEDP 60  
Db 176 LPPTHVQWVPVNFHCRPRRVLRSGLTFSTPTVLPSSGCGSGERTGG----- 225

QY 61 SSGVLRPAAPGGIANPEKKNVC--TQCPNSQSLSS 95  
Db 226 -GGGVTTGAGGIV-FGEACRCGAAHAPTNASMRS 260

RESULT 6  
S6852  
hypothetical protein YOI155c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein O0419; protein AOF1001  
C;Species: Saccharomyces cerevisiae  
C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: S6852; S6854; S67325; S70380  
R;Arino, J.; Casamayor, A.; Gano, F.J.; Gancedo, C.; Lafuente, M.J.; Aldea, M.; Casas,  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S6852  
A;Accession: S6852  
A;Molecule type: DNA  
A;Residues: 1-967 <ARI>  
A;Cross-references: UNIPROT:Q08294; EMBL:Z74897; NID:g1420063; PID:e251930; PID:g142006  
A;Experimental source: strain S288C

A;Molecule type: DNA  
A;Residues: 1-302, 'A', 304-767, 'A', 769-998, 1003-1007, 'S', 1009-1015, 'Q', 1017-1018, 'P', 1020-1029  
A;Cross-references: EMBL:X94215; NID:g1122900; PIDN:CAA63906.1; PID:e214033; PID:g1122900  
A;Experimental source: strain ENY.WA-1A  
C;Genetics:  
A;Gene: SGD:CAT8; MSP8  
A;Cross-references: SGD:S0004893; MIPS:YMR280C  
A;Map position: 13R  
C;Superfamily: GAL4 zinc binuclear cluster homology  
C;Keywords: transmembrane protein  
F;65-102/Domain: GAL4 zinc binuclear cluster homology <GAL4>  
F;456-472/Domain: transmembrane #status predicted <TM1>  
F;738-754/Domain: transmembrane #status predicted <TM2>

Query Match 10.8%; Score 83; DB 2; Length 1433;  
Best Local Similarity 29.4%; Pred. No. 20;  
Matches 32; Conservative 15; Mismatches 50; Indels 12; Gaps 4;

QY 36 LPSAAATDSGPEAGLHAGVLEDPSSNGVLRPAAPGGIANPEKKNVC-GTQCPNSQSLSS 95  
Db 927 LPPATTTSLKPLFGSQSKSLNRQRTPNVKR-----ENPEHYLYGNDNNNNSEA 979

QY 96 G--PLTKQKGLWTTEAKRDAKMSAREVAISVTENIR--QMDRSKVT 140  
Db 980 GHSPMTNTTNGKRLKYEKDAKR-NAKGGIGSGENAHNFQNDTKKMS 1027

RESULT 5  
T37073  
hypothetical protein SCJ30.06c - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T37073  
R;Sanders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999  
A;Reference number: Z21621  
A;Accession: T37073  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-269 <SAN>  
A;Cross-references: UNIPROT:Q9S1Y3; EMBL:AL109973; PIDN:CAB53301.1; GSPDB:GN00070; SCOE  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCJ30.06c

Query Match 10.6%; Score 81; DB 2; Length 269;  
Best Local Similarity 26.8%; Pred. No. 4.6;  
Matches 26; Conservative 9; Mismatches 38; Indels 24; Gaps 4;

QY 11 IEPRYVESWT-----RETSTWLTYSDALPSAAATDSGPEAGLHAGVLEDP 60  
Db 176 LPPTHVQWVPVNFHCRPRRVLRSGLTFSTPTVLPSSGCGSGERTGG----- 225

QY 61 SSGVLRPAAPGGIANPEKKNVC--TQCPNSQSLSS 95  
Db 226 -GGGVTTGAGGIV-FGEACRCGAAHAPTNASMRS 260

RESULT 6  
S6852  
hypothetical protein YOI155c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein O0419; protein AOF1001  
C;Species: Saccharomyces cerevisiae  
C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: S6852; S6854; S67325; S70380  
R;Arino, J.; Casamayor, A.; Gano, F.J.; Gancedo, C.; Lafuente, M.J.; Aldea, M.; Casas,  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S6852  
A;Accession: S6852  
A;Molecule type: DNA  
A;Residues: 1-967 <ARI>  
A;Cross-references: UNIPROT:Q08294; EMBL:Z74897; NID:g1420063; PID:e251930; PID:g142006  
A;Experimental source: strain S288C







Db 489 TTTPEKALSGAGTVAGAVAGSGSGEGAGTEGKNAGVGLGVLASIANPHOKL--- 545

QY 85 TQCPNSQSLSSGPLTKQKQGLWTTTEAKRDAKMSAREVAI 124

Db 546 -----AKRRQL--LEAKRE--RKAAGTLAI 566

RESULT 13

E84824

hypothetical protein At2g40040 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004

C:Accession: E84824

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: E84824

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-839 <STO>

A:Cross-references: UNIPROT:004207; GB:AE002093; NID:G2088657; PIDN:AAB95289.1; GSPDB:GN

C:Genetics:

A:Gene: At2g40040

A:Map position: 2

Query Match 9.9%; Score 76; DB 2; Length 839;

Best Local Similarity 23.8%; Pred. No. 49;

Matches 40; Conservative 14; Mismatches 62; Indels 52; Gaps 6;

QY 5 GSRADAIEPRYVESWTR---ETESTWLTYSDALPSAAA----- 41

Db 338 GSGAGVLGP-----WNKKSSETSGATWGSSDKTKSGAAWNSWKNKIETDSEPAWG 392

QY 42 -----TDSGPEAGGL---HAGVLEDGPGSSNGV-----LRPAAPGGIANPEKKM 81

Db 393 SQGKKNSETSGPAAGAWDKKSKSETEPGAGWGDKKNSHETELGPAAGWNWDKKSDT 452

QY 82 NCGTQCPNSQSLSSGPLTKQKQGLWTTTEAKRDAKMSAREVAISVTEN 129

Db 453 KSGPAAWGSTDAAAGSSDKKN-----SETSDAAAGSRNKKTSIES 496

RESULT 14

S22571

integrase-like protein FB65 - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 31-Dec-1993 #sequence\_revision 02-Aug-1994 #text\_change 09-Jul-2004

C:Accession: S22571; S22572

R:Duilio, A.; Zambrano, N.; Mogavero, A.R.; Ammendola, R.; Cimino, F.; Russo, T. Nucleic Acids Res. 19, S269-S274, 1991

A:Title: A rat brain mRNA encoding a transcriptional activator homologous to the DNA bin

A:Reference number: S22571; MUID:92020215; PMID:1923810

A:Accession: S22571

A:Molecule type: mRNA

A:Residues: 1-499 <DU11>

A:Cross-references: UNIPROT:Q99WK3; EMBL:X60469; NID:G57559; PIDN:CAA42999.1; PID:G57560

A:Accession: S22572

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 31-318 <DU12>

A:Cross-references: EMBL:X60468; NID:G57561; PIDN:CAA42998.1; PID:g1177617

A:Note: this sequence was submitted to the EMBL Data Library, July 1991

C:Genetics:

A:Introns: 88/3; 107/3; 135/3; 157/3; 207/3; 250/2; 252/2; 290/3

C:Keywords: transcription regulation

F:42-78/Domain: WW repeat homology <WW1>

Query Match 9.8%; Score 75.5; DB 2; Length 499;

Best Local Similarity 25.6%; Pred. No. 30;

Matches 34; Conservative 17; Mismatches 59; Indels 23; Gaps 5;

QY 13 PRYIESWTETESTWLTYSDALPSAAATDSGPAGGLHAGVLEDGPGSSNGVLRPAAPG 72

Db 77 PSQGNPSQESQLTWTFG-----AHQGFPEGEF-----WKDEPSEAPMEL----- 118

QY 73 GIANPEKKNCGTQCPNSQSLSSGPLTKQKQGLWTTTEAKRDAKMSAREVA-ISTVTENIR 131

Db 119 GKQPEE---GTLPPFSAQSLSPEVPQEEENLPORNANPGIKCFAVRSGLGWVTEEL 174

QY 132 QMDRSKRVTNKC 144

Db 175 AFGRSSVAVNCCI 187

RESULT 15

T03743

bifocal protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004

C:Accession: T03743

R:Bahri, S.M.; Yang, X.Y.; Chia, W. Mol. Cell. Biol. 17, 5521-5529, 1997

A:Title: The Drosophila bifocal gene encodes a novel protein which colocalizes with activ

A:Reference number: Z15048; MUID:97415628; PMID:92711427

A:Accession: T03743

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1063 <BAH>

A:Cross-references: UNIPROT:016125; EMBL:AF011791; NID:G2388667; PIDN:AAB69991.1; PID:G2

C:Genetics:

A:Cross-references: FlyBase:FBgn0014133

A:Note: bifocal

Query Match 9.8%; Score 75.5; DB 2; Length 1063;

Best Local Similarity 20.7%; Pred. No. 71;

Matches 39; Conservative 23; Mismatches 59; Indels 67; Gaps 6;

QY 5 GSRADAIEPRYVESWTR-----ETESTWLTYSDAL-----PSAAAT- 42

Db 83 GAIADFTPEPATISSTQKRNMGSEKSEKSIINTNSDSTGGHHSVVAVSLSPDAAAT 142

QY 43 -----DSGPAGGLHAGVLEDGPGSSNGVLRPAAPG 72

Db 143 NVTVTIPKQKSSLLNTRSQERENVRYLSSGERDGESEGEQAGVTVNSRCGEVETG 202

QY 73 GIANPEKKNCGTQCPNSQSLSSGPLTKQKQGLWTTTEAKRDAKMSAREVAISVTENIRQ 132

Db 203 TIGSPSSAN---QNPNPHLK---TKCKPGQSVAEKPSAKE-----TIVDNSKS 247

QY 133 MDRSKRVT 140

Db 248 CSKTKSIS 255

Search completed: November 17, 2004, 15:28:51

Job time : 29.6667 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 14:59:59 ; Search time 103.667 Seconds  
(without alignments)  
804.784 Million cell updates/sec

Title: US-10-705-716A-2

Perfect score: 767

Sequence: 1 MCGGSRADAIEPRYYESWT.....VTENIQMDRSKRVTKNCIN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: UniProt\_02.\*

2: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	767	100.0	145	2	Q920K5	Q920k5 rattus norv
2	746	97.3	145	2	Q8VHV1	Q8vhv1 mus musculus
3	645	84.1	145	2	Q9HA93	Q9ha93 homo sapien
4	634	82.7	145	2	Q8WNE9	Q8wne9 sus scrofa
5	617.5	80.5	180	2	Q8WXS3	Q8wxs3 homo sapien
6	473.5	61.7	149	2	Q8WXS1	Q8wxs1 homo sapien
7	459	59.8	123	2	Q8CYS9	Q8cys9 mus musculus
8	294	38.3	54	2	Q790N3	Q790n3 rattus norv
9	294	38.3	54	2	Q8VBS8	Q8vbs8 mus musculus
10	288	37.5	73	2	Q8WXS0	Q8wxs0 homo sapien
11	278	36.2	54	2	Q8WTP6	Q8wtp6 homo sapien
12	278	36.2	54	2	AAH35038	Aah35038 homo sapi
13	272	35.5	80	2	Q8WXS2	Q8wxs2 homo sapien
14	267	34.8	54	2	Q8WNE8	Q8wne8 sus scrofa
15	161	21.0	32	2	Q801V5	Q801v5 brachydanio
16	92.5	12.1	613	2	Q6ZMJ6	Q6zjm6 homo sapien
17	92.5	12.1	613	2	BAD18729	Bad18729 homo sapi
18	92.5	12.1	753	1	CNO3 HUMAN	Cno3515 homo sapien
19	92.5	12.1	1073	1	PVDA PLAKN	P22545 plasmodium
20	89.5	11.7	178	2	Q8W0A0	Q8w0a0 oryza sativ
21	88	11.5	365	1	POLG SUMYS	P25242 sugarcane m
22	87	11.3	1035	2	Q76C74	Q76c74 saccharomyc
23	87	11.3	1035	2	BAD06577	Bad06577 saccharom
24	87	11.3	1713	2	Q8TGB1	Q8tgel saccharomyc
25	87	11.3	1713	2	BAD06576	Bad06576 saccharom
26	86	11.2	585	2	Q7UZ29	Q7uz29 rhodopirell
27	86	11.2	1070	1	PVDG PLAKN	P50494 plasmodium
28	84.5	11.0	534	2	Q96SA2	Q96sa2 homo sapien
29	84	11.0	155	2	Q6U5F5	Q6u5f5 plasmodium
30	84	11.0	155	2	AAQ92950	Aaq92950 plasmodiu
31	83.5	10.9	214	2	Q8MTC2	Q8mtc2 leucophaea

#### ALIGNMENTS

##### RESULT 1

Q920K5 PRELIMINARY; PRT; 145 AA.  
AC Q920K5  
DT 01-DEC-2001 (TREMELrel. 19, Created)  
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)  
DE Dem-A20-4 (BAALC isoform 1-6-8).  
GN Name=dem-A20-4; Synonym=BaalC;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wang X., Tian Q., Li W., Okano A., Suzuki T.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=21574584; PubMed=11707601;  
RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
RA Mrozek K., Sill H., Knuttila S., Kolitz J.E., Archer K.J.,  
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
lineage, is implicated in hematopoiesis and acute leukemia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).  
DR EMBL; AB073318; BAB70507.1; -;  
DR EMBL; AF371321; AAL50517.1; -;  
DR InterPro; IPR009728; BAALC N.  
DR Pfam; PF06989; BAALC N; 1\_-  
SQ SEQUENCE 145 AA; 15475 MW; DSA27AD67456F341 CRC64;

Query Match 100.0%; Score 767; DB 2; Length 145;  
Best Local Similarity 100.0%; Pred. No. 3.4e-61;  
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCGGSRADAIEPRYYESWTRETESTWLTDSALPSAAATDSGPAGGLHAGVLEDP 60  
Db 1 MCGGSRADAIEPRYYESWTRETESTWLTDSALPSAAATDSGPAGGLHAGVLEDP 60  
Qy 61 SSGVLRPAAPGGIANPEKKNMGCTCPNSQSLSSGGLTKQNGLWTEAKRDKRMSAR 120  
Db 61 SSGVLRPAAPGGIANPEKKNMGCTCPNSQSLSSGGLTKQNGLWTEAKRDKRMSAR 120  
Qy 121 EVAISVTENIQMDRSKRVTKNCIN 145  
Db 121 EVAISVTENIQMDRSKRVTKNCIN 145

##### RESULT 2

Q8VHV1 PRELIMINARY; PRT; 145 AA.  
AC Q8VHV1



RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=21574584; PubMed=11707601;  
 RX Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
 RA Mrozek K., Sill H., Knuutila S., Koltz J.E., Archer K.J.,  
 RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 RT lineage, is implicated in hematopoiesis and acute leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RC MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Richards S., Worley K.C., Hale S.G., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywicki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RC Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.  
 DR EMBL; AK022077; BAB13960.1; -;  
 DR EMBL; AF371319; AAL50515.1; -;  
 DR EMBL; BC011517; AAH11517.1; -;  
 DR EMBL; AF363578; AAL50377.1; -;  
 DR InterPro; IPR009728; BAALC N.  
 DR Pfam; PF06989; BAALC N; 1.  
 SQ SEQUENCE 145 AA; 15551 MW; CFB992BBE283DD92E CRC64;  
 Query Match 84.1%; Score 645; DB 2; Length 145;  
 Best Local Similarity 83.4%; Pred. No. 3.3e-50;  
 Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MCGGSRADALEPRYYSWTRTESTTWLTYSALPSAAATDSGPEAGGLHAGVLEDPG 60  
 Db 1 MCGGSRADALEPRYYSWTRTESTTWLTYSALPSAAATDSGPEAGGLHAGVLEDPG 60  
 QY 61 SSGVLRPAAPGGIANPEKKNNCGTCNPSSLSGGLTKQNGLWTTTEAKRDARMSAR 120  
 Db 61 PSNGVPRSTAGGIPNPEKKNTCTCCNPQSLSGGLTKQNGLWTTTEAKRDARMPAK 120  
 QY 121 EVAISVTENIRQMDRSKRVTKNCIN 145  
 Db 121 EVTINVTDISIQMDRSRRITKNCVN 145

RESULT 4  
 Q8WNE9

ID Q8WNE9 PRELIMINARY; PRT; 145 AA.  
 AC Q8WNE9;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE BAALC isoform 1-6-8.  
 GN Name=BAALC;  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=21574584; PubMed=11707601;  
 RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
 RA Mrozek K., Sill H., Knuutila S., Koltz J.E., Archer K.J.,  
 RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 RT lineage, is implicated in hematopoiesis and acute leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).  
 DR EMBL; AF371322; AAL50518.1; -;  
 DR InterPro; IPR009728; BAALC N.  
 DR Pfam; PF06989; BAALC N; 1.  
 SQ SEQUENCE 145 AA; 15401 MW; C75ED7D00EF82E26 CRC64;  
 Query Match 82.7%; Score 634; DB 2; Length 145;  
 Best Local Similarity 82.8%; Pred. No. 3.2e-49;  
 Matches 120; Conservative 10; Mismatches 15; Indels 0; Gaps 0;  
 QY 1 MCGGSRADALEPRYYSWTRTESTTWLTYSALPSAAATDSGPEAGGLHAGVLEDPG 60  
 Db 1 MCGGSRADALEPRYYSWTRTESTTWLTYSALPSAAATDSGPEAGGLHAGVLEDPG 60  
 QY 61 SSGVLRPAAPGGIANPEKKNNCGTCNPSSLSGGLTKQNGLWTTTEAKRDARMSAR 120  
 Db 61 SANGVPRSTAGGTSNPEKKNSCGTCNPQSLSGGLTKQNGLWTTTEAKRDARMSAR 120  
 QY 121 EVAISVTENIRQMDRSKRVTKNCIN 145  
 Db 121 EVTINVTDISIQMDRSRRITKNCIN 145  
 RESULT 5  
 Q8WXS3 PRELIMINARY; PRT; 180 AA.  
 ID Q8WXS3;  
 AC Q8WXS3;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE BAALC 1-5-6-8.  
 GN Name=BAALC;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=21574584; PubMed=11707601;  
 RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
 RA Mrozek K., Sill H., Knuutila S., Koltz J.E., Archer K.J.,  
 RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 RT lineage, is implicated in hematopoiesis and acute leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).  
 DR EMBL; AF363578; AAL50379.1; -;  
 DR InterPro; IPR009728; BAALC N.  
 DR Pfam; PF06989; BAALC N; 1.  
 SQ SEQUENCE 180 AA; 19224 MW; 380183E0F188F684 CRC64;  
 Query Match 80.5%; Score 617.5; DB 2; Length 180;  
 Best Local Similarity 67.2%; Pred. No. 1.3e-47;  
 Matches 121; Conservative 10; Mismatches 14; Indels 35; Gaps 1;







REP	SEQUENCE FROM N.A.
TISU	Brain;
RRA	Strausberg R.;
EMBL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC035038.1; -
SEQ	SEQUENCE 54 AA; 5663 MW; PEF0B4EABED9B829 CRC64;

  

Query Match	36.2%; Score 278; DB 2; Length 54;
Best Local Similarity	94.4%; Pred. No. 1.1e-17;
Matches 51; Conservative	1; Mismatches 2; Indels 0; Gaps

  

QY	1	MCGGSRADAIEPRYESWTRETESTTWLTYSDALPSAAATDSGPAGGLHAG	54
DB	1	MCGGSRADAIEPRYESWTRETESTTWLTYSDALPSAAAPDSGPAGGLHSG	54

  

RESULT 13	
ID Q8WXS2	PRELIMINARY; PRT; 80 AA.
AC Q8WXS2	
DT 01-MAR-2002	(TrEMBLrel. 20, Created)
DT 01-MAR-2002	(TrEMBLrel. 20, Last sequence update)
DT 01-WAR-2004	(TrEMBLrel. 26, Last annotation update)
DE BAALC 1-2.	
GN Name=BAALC;	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RP [1]	
RN SEQUENCE FROM N.A.	
RP MEDLINE=21574584; PubMed=11707601;	
TX Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,	
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,	
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;	
RT "BAALC, the human member of a novel mammalian neuroectoderm gene	
RL lineage, is implicated in hematopoiesis and acute leukemia.";	
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).	
DR EMBL; AF363578; AAL50382.1; -.	
DR InterPro; IPR009728; BAALC_N.	
DR Pfam; PF06989; BAALC_N; 1.	
SEQ	SEQUENCE 80 AA; 8573 MW; 70BE9B91C3245E2D CRC64;

  

Query Match	35.5%; Score 272; DB 2; Length 80;
Best Local Similarity	94.3%; Pred. No. 5.8e-17;
Matches 50; Conservative	1; Mismatches 2; Indels 0; Gaps

  

QY	1	MCGGSRADAIEPRYESWTRETESTTWLTYSDALPSAAATDSGPAGGLHA	53
DB	1	MCGGSRADAIEPRYESWTRETESTTWLTYSDALPSAAAPDSGPAGGLHS	53

  

RESULT 14	
Q8WNE8	PRELIMINARY; PRT; 54 AA.
ID Q8WNE8	
AC Q8WNE8	
DT 01-MAR-2002	(TrEMBLrel. 20, Created)
DT 01-MAR-2002	(TrEMBLrel. 20, Last sequence update)
DT 01-WAR-2004	(TrEMBLrel. 26, Last annotation update)
DE BAALC isoform 1-8.	
GN Name=BAALC;	
OS Sus scrofa (Pig).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.	
OX NCBI_TaxID=9823;	
RP [1]	
RN SEQUENCE FROM N.A.	
RP MEDLINE=21574584; PubMed=11707601;	
TX Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,	
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,	
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;	
RT "BAALC, the human member of a novel mammalian neuroectoderm gene	
RL lineage, is implicated in hematopoiesis and acute leukemia.";	

RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906 (2001).  
DR EMBL; AF371326; AAL50522.1; -.  
DR InterPro; IPR009728; BAALC\_N.  
DR Pfam; PF06989; BAALC\_N; 1.  
SQ SEQUENCE 54 AA; 5681 MW; 6538C3DABED9B825 CRC64;

Query Match 34.8%; Score 267; DB 2; Length 54;  
Best Local Similarity 92.6%; Pred. No. 1e-16;  
Matches 50; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTITDSDALPSAAATDSDGPEAGGLHAG 54  
DB 1 MCGGSRADAEIPRYESWTRETESTWLTITDSDAPPNAPDSDGPEAGGLQAG 54

## RESULT 15

Q801V5 PRELIMINARY; PRT; 32 AA.

AC Q801V5;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE SI:zC215113.3 (Novel protein similar to human brain and acute  
leukemia, cytoplasmic (BAALC)) (Fragment).  
GN Name=SI:zC215113.3;  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_taxid=7955;  
RN [1]

RN SEQUENCE FROM N.A.

RA Sehra H.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL807244; CAD87801.1; .

DR InterPro; IPR009728; BAALC\_N.

DR Pfam; PF06989; BAALC\_N; 1.

FT NON TER 32

SQ SEQUENCE 32 AA; 3663 MW; 916A5445D263B7F7 CRC64;

Query Match 21.0%; Score 161; DB 2; Length 32;  
Best Local Similarity 90.6%; Pred. No. 1.9e-07;  
Matches 29; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCGGSRADAEIPRYESWTRETESTWLTITD 32  
DB 1 MCGGSRADAEIPRYESWTRETESTWLTITNTE 32

Search completed: November 17, 2004, 15:27:45  
Job time : 113.667 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2004, 14:59:59 ; Search time 103.667 seconds  
(without alignments)  
804.784 Million cell updates/sec

Title: US-10-705-716A-4

Perfect score: 78

Sequence: 1 MCGGSRADAEIPRYESWT.....VTDSIQQDRSRITKNCVN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	778	100.0	145	2	Q9HA93	Q9ha93 homo sapien
2	750.5	96.5	180	2	Q8WXS3	Q8wxs3 homo sapien
3	664	85.3	145	2	Q8WNE9	Q8wne9 sus scrofa
4	649	83.4	145	2	Q8VHV1	Q8vhv1 mus musculus
5	645	82.9	145	2	Q920K5	Q920k5 rattus norv
6	569.5	73.2	149	2	Q8WXS1	Q8wxs1 homo sapien
7	381	49.0	123	2	Q9CYS9	Q9cys9 mus musculus
8	309	39.7	73	2	Q8WXS0	Q8wxs0 homo sapien
9	299	38.4	54	2	Q8WTP6	Q8wtp6 homo sapien
10	299	38.4	54	2	AAH35038	Aah35038 homo sapi
11	293	37.7	80	2	Q8WXS2	Q8wxs2 homo sapien
12	282	36.2	54	2	Q8WNE8	Q8wne8 sus scrofa
13	278	35.7	54	2	Q790N3	Q790n3 rattus norv
14	278	35.7	54	2	Q8VB88	Q8vbs8 mus musculus
15	161	20.7	32	2	Q801V5	Q801v5 brachydanio
16	97.5	12.5	306	2	Q72H02	Q72h02 thermus the
17	97.5	12.5	306	2	AA81776	Aa81776 thermus t
18	95.5	12.3	1097	1	S24C ARATH	Q9m291 arabidopsis
19	94.5	12.1	1625	2	O55597	O55597 garlic viru
20	93	12.0	291	2	Q6C2R4	Q6c2r4 candida gla
21	92.5	11.9	717	2	Q6FKP2	Q6fkp2 rattus norv
22	92.5	11.9	1001	1	PSPA RAT	Q9jmc1 rattus norv
23	92	11.8	221	2	Q8GAN3	Q8gan3 arthrobacte
24	91.5	11.8	222	2	Q8LE41	Q8le41 arabidopsis
25	91.5	11.8	249	2	Q9ZX85	Q9zx85 bacterioph
26	91.5	11.8	2042	2	Q767L8	Q767l8 sus scrofa
27	91.5	11.8	2042	2	BAD08434	Bad08434 sus scrof
28	91	11.7	189	2	Q73W33	Q73w33 mycobacteri
29	91	11.7	189	2	AA505144	Aa505144 mycobacte
30	91	11.7	670	2	Q828T2	Q828t2 streptomyc
31	90.5	11.6	330	2	Q6BXP9	Q6bxp9 debaryomyce

32 90 11.6 219 2 Q82AY5 Q82ay5 streptomyc

33 89 11.4 321 2 Q9VUD8 Q9vud8 drosophila

34 89 11.4 629 2 Q87V71 Q87v71 pseudomonas

35 89 11.4 805 2 Q758M4 Q758m4 ashbya goss

36 89 11.4 805 2 AAS52422 Aas52422 ashbya go

37 89 11.4 1064 2 Q8DIP1 Q8dip1 synecococc

38 88.5 11.4 340 2 Q96I20 Q96i20 homo sapien

39 88.5 11.4 690 2 Q9FX17 Q9fx17 arabidopsis

40 88 11.3 302 2 Q93Z47 Q93z47 arabidopsis

41 88 11.3 472 2 Q8UVC3 Q8uvc3 gallus gall

42 88 11.3 472 2 Q8UWC8 Q8uwc8 gallus gall

43 88 11.3 1194 2 Q6C4Z0 Q6c4z0 yarrowia li

44 87.5 11.2 222 2 Q93V72 Q93v72 arabidopsis

45 87.5 11.2 340 2 Q6FHY9 Q6fhy9 homo sapien

## ALIGNMENTS

RESULT 1

Q9HA93 PRELIMINARY; PRT; 145 AA.

ID Q9HA93

AC Q9HA93;

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE Hypothetical protein FLJ12015 (BAALC isoform 1-6-8) (Brain and acute leukemia, cytoplasmic) (BAALC 1-6-8).

DE Name=BAALC;

GN Homo sapiens (Human)

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606;

OX [1]

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Whole embryo;

RX PubMed=14702039;

RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S., Yoshioka K., Matsunawa H., Ichihara T., Shiohata N., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki H., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M., Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K., Mizushima-Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T., Mizushima-Togashi T., Satoh T., Shirai Y., Takahashi Y., Nakagawa K., Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R., Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.

RT \*Complete sequencing and characterization of 21,243 full-length human cDNAs.

RN Nat. Genet. 36:40-45(2004).

RP [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21574584; PubMed=11707601;

RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K., Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,

RA Calligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 lineage, is implicated in hematopoiesis and acute leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906 (2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR ENBL; AK022077; BAB13960.1; -  
 DR ENBL; AF371319; AAL50515.1; -  
 DR ENBL; BC011517; AAH11517.1; -  
 DR ENBL; AF363578; AAL50377.1; -  
 DR InterPro: IPR009728; BAALC\_N.  
 DR Pfam: PF06989; BAALC\_N; 1.  
 DR SEQUENCE 145 AA; 15551 MW; CFB92BBE283DD92E CRC64;  
 Query Match 100.0%; Score 778; DB 2; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-55;  
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIPRYESWTRETESTWLTYSDDAPPSAAPPDGGPAGLHSGMLEDGL 60  
 DB 1 MCGGSRADAIPRYESWTRETESTWLTYSDDAPPSAAPPDGGPAGLHSGMLEDGL 60  
 QY 61 PSNGVPRSTAPGGIPNPEKKNCTCQCPNQSLSGGPLTKQNGLOTTAKRDAKMPAK 120  
 DB 61 PSNGVPRSTAPGGIPNPEKKNCTCQCPNQSLSGGPLTKQNGLOTTAKRDAKMPAK 120  
 QY 121 EVTINVTDSIQMDRSRRITKNCVN 145  
 DB 121 EVTINVTDSIQMDRSRRITKNCVN 145  
 RESULT 2  
 Q8WXS3 PRELIMINARY; PRT; 180 AA.  
 AC Q8WXS3;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE BAALC 1-5-6-8.  
 GN Name=BAALC;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=21574584; PubMed=11707601;  
 RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
 RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,  
 RA Calligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 lineage, is implicated in hematopoiesis and acute leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906 (2001).  
 DR ENBL; AF363578; AAL50379.1; -  
 DR InterPro: IPR009728; BAALC\_N.  
 DR Pfam: PF06989; BAALC\_N; 1.  
 DR SEQUENCE 180 AA; 19224 MW; 380183E0F188F684 CRC64;  
 Query Match 96.5%; Score 750.5; DB 2; Length 180;  
 Best Local Similarity 80.6%; Pred. No. 7.5e-53;  
 Matches 145; Conservative 0; Mismatches 0; Indels 35; Gaps 1;  
 QY 1 MCGGSRADAIPRYESWTRETESTWLTYSDDAPPSAAPPDGGPAGLHSGMLEDGL 53  
 DB 1 MCGGSRADAIPRYESWTRETESTWLTYSDDAPPSAAPPDGGPAGLHSGMLEDGL 60  
 QY 54 -----GMLEDGLPSNGVPRSTAPGGIPNPEKKNCT 85  
 DB 61 KIKAPTDSVDEGLPSASXMAPLAVFSGHMGLEDGLPSNGVPRSTAPGGIPNPEKKNCT 120  
 QY 86 QCPNPQSLSSGPLTKQNGLOTTAKRDAKMPAKKEVTINVTDSIQMDRSRRITKNCVN 145  
 DB 121 QCPNPQSLSSGPLTKQNGLOTTAKRDAKMPAKKEVTINVTDSIQMDRSRRITKNCVN 180  
 RESULT 3  
 Q8WNE9 PRELIMINARY; PRT; 145 AA.  
 AC Q8WNE9;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE BAALC isoform 1-6-8.  
 GN Name=BAALC;  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21574584; PubMed=11707601;  
 RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
 RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,  
 RA Calligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 lineage, is implicated in hematopoiesis and acute leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906 (2001).  
 DR ENBL; AF371322; AAL50518.1; -  
 DR InterPro: IPR009728; BAALC\_N.  
 DR Pfam: PF06989; BAALC\_N; 1.  
 DR SEQUENCE 145 AA; 15401 MW; C75ED7D00EF82E26 CRC64;  
 Query Match 85.3%; Score 664; DB 2; Length 145;  
 Best Local Similarity 84.1%; Pred. No. 5.5e-46;  
 Matches 122; Conservative 12; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIPRYESWTRETESTWLTYSDDAPPSAAPPDGGPAGLHSGMLEDGL 60  
 DB 1 MCGGSRADAIPRYESWTRETESTWLTYSDDAPPSAAPPDGGPAGLHSGMLEDGL 60  
 QY 61 PSNGVPRSTAPGGIPNPEKKNCTCQCPNQSLSGGPLTKQNGLOTTAKRDAKMPAK 120  
 DB 61 PSNGVPRSTAPGGIPNPEKKNCTCQCPNQSLSGGPLTKQNGLOTTAKRDAKMPAK 120  
 QY 121 EVTINVTDSIQMDRSRRITKNCVN 145  
 DB 121 EVTINVTDSIQMDRSRRITKNCVN 145

```

RESULT 4
Q8VHV1 PRELIMINARY; PRT; 145 AA.
AC Q8VHV1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE BAALC isoform 1-6-8 (Mus musculus 16 days neonate cerebellum cDNA,
DE RIKEN full-length enriched library, clone:9630028H16 product:brain and
DE acute leukemia, cytoplasmic, full insert sequence).
GN Name=BaalC;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
EN [1]
EN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=21574584; PubMed=11707601;
RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;
RT "BAALC, the human member of a novel mammalian neuroectoderm gene
RT lineage, is implicated in hematopoiesis and acute leukemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).
EN [2]
EN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
EN [3]
EN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
EN [4]
EN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
EN [5]
EN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Katsunai T., Tashiro H., Itoh M.,
RA Konno H., Akiyama J., Nishi K., Kitamura S., Hazama M., Nishine T., Harada A.,
RA Sumi N., Tahii Y., Nakamura S., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Okawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
EN [7]
EN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;

```

```

RA Adachi J., Aizawa K., Akimura T., Atakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tomaru A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF371320; AAL50516.1; -
DR EMBL; AK079337; BAC37611.1; -
DR MGD; MGI:1928704; BaalC.
DR InterPro; IPR009728; BAALC_N.
DR Pfam; PF06989; BAALC_N; 1.
SQ SEQUENCE 145 AA; 15515 MW; 4972670A618C4D6D CRC64;

Query Match 83.4%; Score 649; DB 2; Length 145;
Best Local Similarity 83.4%; Pred. No. 8.9e-45;
Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 MCGCGSRADAIEPRYIESWTRETESTWLTYSDDAPPSSAAAPDSGPEAGLHSGMLEDGL 60
Db 1 MCGCGSRADAIEPRYIESWTRETESTWLTYSDDAPPSSAAAPDSGPEAGLHSGMLEDGL 60

QY 61 PSNGVPRSTAGGIPNPKKTCNCPNPSLSSGPTOKNGLOTTTEAKDKRMPAK 120
Db 61 SSSGVLPPAAGGIANPEKQNCCTQCPSNQLSSGPTOKNGLOTTTEAKDKRMSAR 120

QY 121 EVTINVTDISIQQMDRSRRITKNCVN 145
Db 121 EVAINVTENIQMDRSRRITKNCIN 145

RESULT 5
Q920K5 PRELIMINARY; PRT; 145 AA.
AC Q920K5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Dem-A20-4 (BAALC isoform 1-6-8).
GN Name=dem-A20-4; Synonyms=BaalC;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
EN [1]
EN SEQUENCE FROM N.A.
RP Wang X., Tian Q., Li W., Okano A., Suzuki T.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=21574584; PubMed=11707601;
RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;
RT "BAALC, the human member of a novel mammalian neuroectoderm gene
RT lineage, is implicated in hematopoiesis and acute leukemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).
DR EMBL; AB073318; BAB70507.1; -
DR EMBL; AF371321; AAL50517.1; -
DR InterPro; IPR009728; BAALC_N.
DR Pfam; PF06989; BAALC_N; 1.
SQ SEQUENCE 145 AA; 15475 MW; DSA27AD67456F341 CRC64;

Query Match 82.9%; Score 645; DB 2; Length 145;
Best Local Similarity 83.4%; Pred. No. 1.9e-44;
Matches 121; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

```

```
QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDDAPPSAAAPDSGPGAGGLHSGMLDGL 60
Db 1 MCGGSRADAEIPRYESWTRETESTWLTYSDDAPPSAAATDSGPGAGGLHAGVLEDGP 60
QY 61 PSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSSGGLPTQKQGLQTEAKRAKMPAK 120
Db 61 SSNGVLRPAAPGGIANPEKKNKCTQCPNSQSSGGLPTQKQGLQTEAKRAKMSAR 120
QY 121 EVTINVTDSIQMDRSRRITKNCVN 145
Db 121 EVAISVTENRQMDRSKRVTKNCIN 145
RESULT 6
Q8WXS1 PRELIMINARY; PRT; 149 AA.
AC Q8WXS1
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE BAALC 1-5-6-7-8.
GN Name=BAALC;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21574584; PubMed=11707601;
RA Tanner S.M., Austin J.B., Leone G., Rush L.J., Plass C., Heinonen K.,
RA Mrozek K., Sill H., Knuttila S., Kolitz J.E., Archer K.J.,
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;
RT "BAALC, the human member of a novel mammalian neuroectoderm gene
RT lineage, is implicated in hematopoiesis and acute leukemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).
DR EMBL; AF363578; AAU50381.1; -
DR Genbank; HGNC:14333; BAALC.
DR InterPro; IPR009728; BAALC_N.
DR Pfam; PF06989; BAALC_N; 1.
SQ SEQUENCE 149 AA; 15577 MW; DB26C19969B91F6E CRC64;
Query Match 73.2%; Score 569.5; DB 2; Length 149;
Best Local Similarity 75.7%; Pred. No. 2.4e-38;
Matches 109; Conservative 0; Mismatches 0; Indels 35; Gaps 1;
QY 1 MCGGSRADAEIPRYESWTRETESTWLTYSDDAPPSAAAPDSGPGAGGLHS----- 53
Db 1 MCGGSRADAEIPRYESWTRETESTWLTYSDDAPPSAAAPDSGPGAGGLHSLVLEAKS 60
QY 54 -----GMLEDGLPSNGVPRSTAPGGIPNPEKTKNCET 85
Db 61 KIKAPTDVSDEGLFSASKMAPLAVFSGHGMLEDGLPSNGVPRSTAPGGIPNPEKTKNCET 120
QY 86 QCPNPQSSGGLPTQKQGLQTE 109
Db 121 QCPNPQSSGGLPTQKQGLQTE 144
RESULT 7
Q9CYS9 PRELIMINARY; PRT; 123 AA.
AC Q9CYS9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length
DE enriched library, clone:2810457D07 product:brain and acute leukemia,
DE cytoplasmic, full insert sequence.
GN Name=BaalC;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=92279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Whole body;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Whole body;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
RA Okazaki Y., Okido T., Owa C., Saito R., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Teijima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR DR EMBL; AK013358; BAE28808.1; -
DR DR MGD; MGI:1928704; Baalc.
SQ SEQUENCE 123 AA; 13391 MW; 33DFCB5F8FCAA0AB CRC64;
Query Match 49.0%; Score 381; DB 2; Length 123;
Best Local Similarity 77.4%; Pred. No. 3e-23;
Matches 72; Conservative 9; Mismatches 12; Indels 0; Gaps 0;
QY 53 SCMLDGLPSNGVPRSTAPGGIPNPEKTKNCETQCPNPQSSGGLPTQKQGLQTEAKR 112
Db 31 SGVLEDGLSSNGVLRPAAPGGIANPEKKNKCTQCPNSQSSGGLPTQKQGLQTEAKR 90
QY 113 DAKRMPAKEVTINVTDSIQMDRSRRITKNCVN 145
```







DR EMBL; AF371324; AAL50520.1; -.  
DR MGD; MGI:1928704; Baalc.  
DR InterPro; IPR009728; BAALC\_N.  
DR Pfam; PF06989; BAALC\_N; 1.  
SQ SEQUENCE 54 AA; 5667 MW; FEE8C0EBFBDCDB829 CRC64;

Query Match 35.7%; Score 278; DB 2; Length 54;  
Best Local Similarity 94.4%; Pred. No. 2.3e-15;  
Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCGGSRADAIETPRYIESWTRETESTWLTYSDDAPPAAAPDSGPEAGLHSG 54  
DB 1 MCGGSRADAIETPRYIESWTRETESTWLTYSDDAPPAAAPDSGPEAGLHAG 54

RESULT 15

Q801V5 PRELIMINARY; PRT; 32 AA.

AC Q801V5;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE SI:zC215113.3 (Novel protein similar to human brain and acute  
DE leukemia, cytoplasmic (BAALC)) (Fragment).  
GN Name=SI:zC215113.3;  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sehra H.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL807244; CAD87801.1; -.  
DR InterPro; IPR009728; BAALC\_N.  
DR Pfam; PF06989; BAALC\_N; 1.  
FT NON TER 32  
SQ SEQUENCE 32 AA; 3663 MW; 916A5445D263B7F7 CRC64;

Query Match 20.7%; Score 161; DB 2; Length 32;  
Best Local Similarity 90.6%; Pred. No. 3.4e-06;  
Matches 29; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MCGGSRADAIETPRYIESWTRETESTWLTYSDDAPPAAAPDSGPEAGLHSG 32  
DB 1 MCGGSRADAIETPRYIESWTRETESTWLTYSDDAPPAAAPDSGPEAGLHAG 32



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: November 17, 2004, 14:59:59 ; Search time 103.667 Seconds  
(without alignments)  
804.784 Million cell updates/sec

Title: US-10-705-716A-8  
Perfect score: 767  
Sequence: 1 MCGGSRADAIEPRYYESWT.....VTENIRQMDRSKRVTKNCIN 145

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 40 summaries

Database : UniProt 02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	767	100.0	145	2 Q8VHV1	Q8VHV1 mus musculus
2	746	97.3	145	2 Q920K5	Q920K5 ractus norv
3	649	84.6	145	2 Q9HA93	Q9HA93 homo sapien
4	634	82.7	145	2 Q8WNE9	Q8WNE9 sus scrofa
5	621.5	81.0	180	2 Q8WXS3	Q8WXS3 homo sapien
6	480	62.6	123	2 Q9CYS9	Q9CYS9 mus musculus
7	472.5	61.6	149	2 Q8WXS1	Q8WXS1 homo sapien
8	294	38.3	54	2 Q790N3	Q790N3 ractus norv
9	294	38.3	54	2 Q8VBS8	Q8VBS8 mus musculus
10	288	37.5	73	2 Q8WXS0	Q8WXS0 homo sapien
11	278	36.2	54	2 Q8WTP6	Q8WTP6 homo sapien
12	278	36.2	54	2 AAH35038	AAH35038 homo sapi
13	272	35.5	80	2 Q8WXS2	Q8WXS2 homo sapien
14	267	34.8	54	2 Q8WNE8	Q8WNE8 sus scrofa
15	161	21.0	32	2 Q801V5	Q801V5 brachydanio
16	87.5	11.4	1073	1 PVDA_PLAKN	P22545 plasmodium
17	86	11.2	1449	2 Q7XPB1	Q7XPB1 oryza sativ
18	85.5	11.1	178	2 Q8W0A0	Q8W0A0 oryza sativ
19	85	11.1	882	1 AREA_ASPNG	Q13412 aspergillus
20	85	11.1	882	2 CAA68196	Caa68196 aspergill
21	84.5	11.0	367	2 Q7K2P9	Q7K2P9 drosophila
22	84.5	11.0	494	2 Q8IR46	Q8IR46 drosophila
23	84.5	11.0	495	2 Q961T1	Q961T1 drosophila
24	84.5	11.0	591	2 Q83H42	Q83H42 tropheryma
25	84.5	11.0	591	2 Q83NY7	Q83NY7 tropheryma
26	84.5	11.0	659	2 Q9Y16	Q9Y16 drosophila
27	84.5	11.0	659	2 AAF48389	Aaf48389 drosophil
28	84.5	11.0	773	2 Q9FNV4	Q9FNV4 porphyra ye
29	84	11.0	253	2 Q66155	Q66155 streptomyce
30	84	11.0	365	1 POLG_SUMVS	P25242 sugarcane m
31	84	11.0	1433	1 CAT8_YEAST	P39113 saccharomyc

32	83.5	10.9	291	2	Q871V5	Q871V5 neurospora
33	83.5	10.9	865	1	NREA_PENUR	Q92269 penicillium
34	83	10.8	476	2	Q7PV51	Q7PV51 anopheles g
35	83	10.8	585	2	Q7UZ29	Q7UZ29 rhodospirill
36	83	10.8	860	1	AREA_PENRO	Q13508 rhodospirill
37	81.5	10.6	214	2	Q8MTC2	Q8MTC2 leucophaea
38	81.5	10.6	1100	2	Q91LA3	Q91LA3 white spot
39	81	10.6	219	2	Q82AV5	Q82AV5 streptomyce
40	81	10.6	228	2	Q91TQ3	Q91TQ3 tupais herp
41	81	10.6	516	2	Q74641	Q74641 aspergillus
42	81	10.6	1070	1	PVDG_PLAKN	P50494 plasmodium
43	80.5	10.5	155	2	Q6U5F5	Q6U5F5 plasmodium
44	80.5	10.5	155	2	AAQ92950	AAQ92950 plasmodiu
45	80.5	10.5	852	1	AOF2_HUMAN	O60341 homo sapien

## ALIGNMENTS

### RESULT 1

ID	Q8VHV1	PRELIMINARY;	PRT;	145 AA.
AC	Q8VHV1;			
DT	01-MAR-2002 (Tremblrel. 20, Created)			
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)			
DT	01-OCT-2004 (Tremblrel. 28, Last annotation update)			
DE	BAALC isoform 1-6-8 (Mus musculus 16 days neonate)			
DE	RIKEN full-length enriched library, clone:9630028H16 product:brain and			
DE	acute leukemia, cytoplasmic, full insert sequence).			
GN	Name=BaalC;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129S6/SvEvTac;			
RX	MEDLINE=21574584; PubMed=11707601;			
RA	Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,			
RA	Mrozek K., Sall H., Knuutila S., Kolitz J.E., Archer K.J.,			
RA	Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;			
RT	"BAALC, the human member of a novel mammalian neuroectoderm gene			
RT	lineage, is implicated in hematopoiesis and acute leukemia.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
RX	MEDLINE=99279253; PubMed=10349636;			
RA	Carninci P., Hayashizaki Y.;			
RT	"High-efficiency full-length cDNA cloning.";			
RL	Meth. Enzymol. 303:19-44(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	RIKEN FANTOM Consortium;			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 403:685-690(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
RA	The FANTOM Consortium,			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RL	60,770 full-length cDNAs.";			
RN	Nature 420:563-573(2002).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
RX	MEDLINE=20499374; PubMed=11042159;			
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,			
RA	Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;			
RT	"Normalization and subtraction of cap-trapper-selected cDNAs to			

RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).

[6]  
 RN SSSEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Nishine T., Harada A.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsubara M.,  
 RA Itoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsubara S., Kawai J.,  
 RA Okazaki Y., Murakatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [7]  
 RN SSSEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF371320; AAL50516.1; -;  
 DR EMBL; AK079337; BAC37611.1; -;  
 DR MGD; MGI:1928704; Baal.  
 DR InterPro: IPR009728; BAALC\_N.  
 DR Pfam: PF06989; BAALC\_N.1.  
 SQ SEQUENCE 145 AA; 15515 MW; 4972670A618C4D6D CRC64;

Query Match 100.0%; Score 767; DB 2; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-64;  
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIERYESWTRETESTWLTYSDDLPSAAATDSGPAGGLHAGVLEDGL 60  
 DB 1 MCGGSRADAIERYESWTRETESTWLTYSDDLPSAAATDSGPAGGLHAGVLEDGL 60  
 QY 61 SSSGVLPRAPAGGIANPEKKNCGTCPCNSQNSLSSGGLTKQKGLWATEAKRDKMSAR 120  
 DB 61 SSSGVLPRAPAGGIANPEKKNCGTCPCNSQNSLSSGGLTKQKGLWATEAKRDKMSAR 120  
 QY 121 EVAINVTENIRQMDRSKRVTKNCIN 145  
 DB 121 EVAINVTENIRQMDRSKRVTKNCIN 145

RESULT 2  
 ID Q920K5 PRELIMINARY; PRT; 145 AA.  
 AC Q920K5  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Dem-A20-4 (BAALC Isoform 1-6-8).  
 GN Name=Dem-A20-4; Synonyms=BaalC;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RN SSSEQUENCE FROM N.A.  
 RA Wang X., Tian O., Li W., Okano A., Suzuki T.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SSSEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=21574584; PubMed=11707601;  
 RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
 RA Mrozek K., Sill H., Knuttila S., Kolitz J.E., Archer K.J.,  
 RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 RT lineage, is implicated in hematopoiesis and acute leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).  
 DR EMBL; AB073318; BAB70507.1; -;  
 DR EMBL; AF371321; AAL50517.1; -;  
 DR InterPro: IPR009728; BAALC\_N.  
 DR Pfam: PF06989; BAALC\_N.1.  
 SQ SEQUENCE 145 AA; 15475 MW; D5A27AD67456F341 CRC64;  
 Query Match 97.3%; Score 746; DB 2; Length 145;  
 Best Local Similarity 97.2%; Pred. No. 2e-62;  
 Matches 141; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIERYESWTRETESTWLTYSDDLPSAAATDSGPAGGLHAGVLEDGL 60  
 DB 1 MCGGSRADAIERYESWTRETESTWLTYSDDLPSAAATDSGPAGGLHAGVLEDGL 60  
 QY 61 SSSGVLPRAPAGGIANPEKKNCGTCPCNSQNSLSSGGLTKQKGLWATEAKRDKMSAR 120  
 DB 61 SSSGVLPRAPAGGIANPEKKNCGTCPCNSQNSLSSGGLTKQKGLWATEAKRDKMSAR 120  
 QY 121 EVAINVTENIRQMDRSKRVTKNCIN 145  
 DB 121 EVAINVTENIRQMDRSKRVTKNCIN 145

RESULT 3  
 Q9HA93  
 ID Q9HA93 PRELIMINARY; PRT; 145 AA.  
 AC Q9HA93  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein FLJ12015 (BAALC isoform 1-6-8) (Brain and acute  
 DE leukemia, cytoplasmic) (BAALC 1-6-8).  
 GN Name=BAALC;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SSSEQUENCE FROM N.A.  
 RC TISSUE=Whole embryo;  
 RX PubMed=14702039;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,  
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
 RA Abe K., Kamiyama K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
 RA Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,  
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Inose N.,  
 RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
 RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,

ID	Q8WNE9	PRELIMINARY;	PRT;	145 AA.
AC	Q8WNE9			
DT	01-WAR-2002	(TRENBLrel. 20, Created)		
DT	01-WAR-2002	(TRENBLrel. 20, Last sequence update)		
DT	01-WAR-2004	(TRENBLrel. 26, Last annotation update)		
DT	01-WAR-2004	(TRENBLrel. 26, Last annotation update)		
DE	BAALC	isoform 1-6-8.		
GN	Name=BAALC;			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa;			
OC	Mammalia; Eutheria;			
OC	Cetartiodactyla; Suina; Suidae; Sus.			
NCBI_TaxID=9823;				
[1]				
SEQUENCE FROM N.A.				
RP	MEDLINE=21574584; PubMed=11707601;			
RA	Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,			
RA	Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,			
RA	Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;			
RT	"BAALC, the human member of a novel mammalian neuroectoderm gene			
RT	lineage, is implicated in hematopoiesis and acute leukemia.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).			
DR	EMBL; AF371322; AAL50518.1; -.			
DR	InterPro; IPR009728; BAALC_N.			
DR	Pfam; PF06989; BAALC_N; 1.			
DR	SEQUENCE 145 AA; 15401 MW; C75ED7D00EF82E26 CRC64;			
Query Match	82.7%;	Score 634;	DB 2;	Length 145;
Best Local Similarity	82.1%;	Pred. No. 7.3e-52;		
Matches 119;	Conservative	11;	Mismatches 15;	Indels 0; Gaps 0;
Qy	1	MCGCGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGGLHAGVLEGL	60	
Db	1	MCGCGSRADAIEPRYYESWTRETESTWLTYSALPSAAATDSGPEAGGLHAGVLEGL	60	
Qy	61	SSNGVLPAAPGGTANPEKKNCGTCQNSONLSGGPLTQKQNGIWAATEAKDKRMSAR	120	
Db	61	SANGVPSATPGGTSNPEKKNCGTCQNSONLSGGPLTQKQNGIWAATEAKDKRMSAR	120	
Qy	121	EVAINVTENIRQMDRSKVTIKNCIN	145	
Db	121	EVTVNVTESIRQVDRNQRIKKCIN	145	
RESULT 5				
Q8WXS3		PRELIMINARY;	PRT;	180 AA.
ID	Q8WXS3			
AC	Q8WXS3			
DT	01-WAR-2002	(TRENBLrel. 20, Created)		
DT	01-WAR-2002	(TRENBLrel. 20, Last sequence update)		
DT	01-WAR-2004	(TRENBLrel. 26, Last annotation update)		
DE	BAALC	1-5-6-8.		
GN	Name=BAALC;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa;			
OC	Mammalia; Eutheria;			
OC	Primates; Catarrhini; Hominoidea; Homo.			
NCBI_TaxID=9606;				
[1]				
SEQUENCE FROM N.A.				
RP	MEDLINE=21574584; PubMed=11707601;			
RA	Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,			
RA	Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,			
RA	Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;			
RT	"BAALC, the human member of a novel mammalian neuroectoderm gene			
RT	lineage, is implicated in hematopoiesis and acute leukemia.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).			
DR	EMBL; AF363578; AAL50379.1; -.			
DR	InterPro; IPR009728; BAALC_N.			
DR	Pfam; PF06989; BAALC_N; 1.			
DR	SEQUENCE 180 AA; 19224 MW; 380183E0F188F684 CRC64;			
Query Match	81.0%;	Score 621.5;	DB 2;	Length 180;
Best Local Similarity	67.2%;	Pred. No. 1.4e-50;		
Matches 121;	Conservative	10;	Mismatches 14;	Indels 35; Gaps 1

```

QY 1 MCGGSRADAIAPRYESWTRETESTWLTYSDDLPSAAATDSQPEAGGLHA----- 53
Db 1 MCGGSRADAIAPRYESWTRETESTWLTYSDDLPSAAATDSQPEAGGLHSVLEAEKS 60
QY 54 -----GVLEDGLSSNGVLRPAAPGGIANPEKKMGCTGCPNSQNLSSGFLTQKQGLWATEAKR 120
Db 61 KIKAPTDSVDSDEGLFSASKMAPLAVFSGHMLDGLPSNGVPRSTAPGGIPNPEKTKNCBT 120
QY 86 QCPNSQNLSSGFLTQKQGLWATEAKRDAKMSAREVAINVTENIRQMDRSKRVTKNCIN 145
Db 121 QCPNPQSLSSGFLTQKQGLWATEAKRDAKMSAREVAINVTENIRQMDRSKRVTKNCIN 180

RESULT 6
Q9CYS9 PRELIMINARY; PRT; 123 AA.
AC Q9CYS9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length
DE enriched library, clone:2810457D07 product:brain and acute leukemia,
DE cytoplasmic, full insert sequence.
GN Name=BaalC;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P.; Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P.; Shibata Y.; Hayatsu N.; Sugahara Y.; Shibata K.; Itoh M.,
RA Konno H.; Okazaki Y.; Muramatsu M.; Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K.; Itoh M.; Aizawa K.; Nagaoka S.; Sasaki N.; Carninci P.,
RA Konno H.; Akiyama J.; Nishi K.; Kitsuai T.; Tashiro H.; Itoh M.,
RA Sumi N.; Ishii Y.; Nakamura S.; Hazama M.; Nishine T.; Harada A.,
RA Yamamoto R.; Matsumoto H.; Sakaguchi S.; Ikegami T.; Kashiwagi K.,
RA Fujiwaka S.; Inoue K.; Togawa Y.; Izawa M.; Ohara E.; Watahiki M.,
RA Yoneda Y.; Ishikawa T.; Ozawa K.; Tanaka T.; Matsuda S.; Kawai J.,
RA Okazaki Y.; Muramatsu M.; Inoue Y.; Kira A.; Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).

```

```

[6]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Whole body;
RC Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Akazawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK013358; BAB28808.1; -.
DR MGD; MGI:1928704; BaalC.
SQ SEQUENCE 123 AA; 13391 MW; 33DFCB5F8CFAA0AB CRC64;

Query Match 62.6%; Score 480; DB 2; Length 123;
Best Local Similarity 98.9%; Pred. No. 2e-37;
Matches 92; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 53 AGVLEDGLSSNGVLRPAAPGGIANPEKKMGCTGCPNSQNLSSGFLTQKQGLWATEAKR 112
Db 31 SGVLEDGLSSNGVLRPAAPGGIANPEKKMGCTGCPNSQNLSSGFLTQKQGLWATEAKR 90

QY 113 DAKMSAREVAINVTENIRQMDRSKRVTKNCIN 145
Db 91 DAKMSAREVAINVTENIRQMDRSKRVTKNCIN 123

RESULT 7
Q8WXS1 PRELIMINARY; PRT; 149 AA.
AC Q8WXS1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE BAALC 1-5-6-7-8.
GN Name=BAALC;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC MEDLINE=21574584; PubMed=11707601;
RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,
RA Calligaris M.A., Bloomfield C.D., de la Chapelle A.;
RT "BAALC, the human member of a novel mammalian neuroectoderm gene
RT lineage, is implicated in hematopoiesis and acute leukemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).
DR EMBL; AF363578; AAL50381.1; -.
DR Genew; HGNC:14333; BAALC_N.
DR InterPro; IPR009728; BAALC_N.
DR Pfam; PF06989; BAALC_N; 1.
SQ SEQUENCE 149 AA; 15577 MW; DB26C19969B91F6E CRC64;

Query Match 61.6%; Score 472.5; DB 2; Length 149;
Best Local Similarity 65.3%; Pred. No. 1.3e-36;
Matches 94; Conservative 3; Mismatches 12; Indels 35; Gaps 1;

QY 1 MCGGSRADAIAPRYESWTRETESTWLTYSDDLPSAAATDSQPEAGGLHA----- 53
Db 1 MCGGSRADAIAPRYESWTRETESTWLTYSDDLPSAAATDSQPEAGGLHSVLEAEKS 60
QY 54 -----GVLEDGLSSNGVLRPAAPGGIANPEKKMGCTGCPNSQNLSSGFLTQKQGLWATEAKR 120
Db 61 KIKAPTDSVDSDEGLFSASKMAPLAVFSGHMLDGLPSNGVPRSTAPGGIPNPEKTKNCBT 120
QY 86 QCPNSQNLSSGFLTQKQGLWATEAKRDAKMSAREVAINVTENIRQMDRSKRVTKNCIN 145
Db 121 QCPNPQSLSSGFLTQKQGLWATEAKRDAKMSAREVAINVTENIRQMDRSKRVTKNCIN 180

```



Db 121 QCPNPQSLSSGPILOKQNGLOTTE 144  
||||| : : : : : ||||| ||

```
RESULT 8
Q790N3 PRELIMINARY; PRT; 54 AA.
AC Q790N3;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE BAALC isoform 1-8.
GN Name=BaalC;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=21574584; PubMed=11707601;
RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;
RT "BAALC, the human member of a novel mammalian neuroectoderm gene
RT lineage, is implicated in hematopoiesis and acute leukemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906 (2001).
DR EMBL; AF371325; AAL50521.1; -.
DR InterPro; IPR009728; BAALC_N.
DR Pfam; PF06989; BAALC_N; 1.
SQ SEQUENCE 54 AA; 5667 MW; FEE8C0EBFBCDB829 CRC64;
```

Query Match 38.3%; Score 294; DB 2; Length 54;  
Best Local Similarity 100.0%; Pred. No. 2.6e-20;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MCGGSRADAIEPRYESWTRETESTWLTYYTDSALPSAAATDSGPAGGLHAG 54
Db 1 MCGGSRADAIEPRYESWTRETESTWLTYYTDSALPSAAATDSGPAGGLHAG 54
```

```
RESULT 9
Q8VBS8 PRELIMINARY; PRT; 54 AA.
AC Q8VBS8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 26, Last annotation update)
DE BAALC isoform 1-8.
GN Name=BaalC;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12986/SvEvTac;
RX MEDLINE=21574584; PubMed=11707601;
RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;
RT "BAALC, the human member of a novel mammalian neuroectoderm gene
RT lineage, is implicated in hematopoiesis and acute leukemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906 (2001).
DR EMBL; AF371324; AAL50520.1; -.
DR MGD; MGI:1928704; BaalC.
DR InterPro; IPR009728; BAALC_N.
DR Pfam; PF06989; BAALC_N; 1.
SQ SEQUENCE 54 AA; 5667 MW; FEE8C0EBFBCDB829 CRC64;
```

Query Match 38.3%; Score 294; DB 2; Length 54;  
Best Local Similarity 100.0%; Pred. No. 2.6e-20;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MCGGSRADAIEPRYESWTRETESTWLTYYTDSALPSAAATDSGPAGGLHAG 54
Db 1 MCGGSRADAIEPRYESWTRETESTWLTYYTDSALPSAAATDSGPAGGLHAG 54
```

```
RESULT 10
Q8WXS0 PRELIMINARY; PRT; 73 AA.
AC Q8WXS0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE BAALC 1-4-5-6-8.
GN Name=BAALC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21574584; PubMed=11707601;
RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;
RT "BAALC, the human member of a novel mammalian neuroectoderm gene
RT lineage, is implicated in hematopoiesis and acute leukemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906 (2001).
DR EMBL; AF363578; AAL50380.1; -.
DR InterPro; IPR009728; BAALC_N.
DR Pfam; PF06989; BAALC_N; 1.
SQ SEQUENCE 73 AA; 7871 MW; 98DBC2E6E6EF524A CRC64;
```

Query Match 37.5%; Score 288; DB 2; Length 73;  
Best Local Similarity 91.4%; Pred. No. 1.4e-19;  
Matches 53; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy 1 MCGGSRADAIEPRYESWTRETESTWLTYYTDSALPSAAATDSGPAGGLHAGVLE 58
Db 1 MCGGSRADAIEPRYESWTRETESTWLTYYTDSAPPSAAAPDSGPAGGLHSGCLEE 58
```

```
RESULT 11
Q8WTP6 PRELIMINARY; PRT; 54 AA.
AC Q8WTP6;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE BAALC isoform 1-8 (BAALC protein) (BAALC 1-8).
GN Name=BAALC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21574584; PubMed=11707601;
RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,
RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,
RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;
RT "BAALC, the human member of a novel mammalian neuroectoderm gene
RT lineage, is implicated in hematopoiesis and acute leukemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
```

RA Stapleton M., Soares M.B., Ronaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF371323; AAL50519.1; -  
 DR EMBL; BC035038; AAH35038.1; -  
 DR EMBL; AF363578; AAL50378.1; -  
 DR InterPro; IPR009728; BAALC\_N.  
 DR Pfam; PF06989; BAALC\_N; 1.  
 SQ SEQUENCE 54 AA; 5663 MW; FEF0B4EABED9B829 CRC64;  
 Query Match 36.2%; Score 278; DB 2; Length 54;  
 Best Local Similarity 94.4%; Pred. No. 8.5e-19;  
 Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIPRYESWTRETESTWLTYSDDLPSAAATDSGPEAGLHAG 54  
 Db 1 MCGGSRADAIPRYESWTRETESTWLTYSDDLPSAAATDSGPEAGLHAG 54  
 RESULT 12  
 AAH35038  
 ID AAH35038 PRELIMINARY; PRT; 54 AA.  
 AC AAH35038;  
 DT 02-MAR-2004 (T-EMBLrel. 27, Created)  
 DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)  
 DT 02-MAR-2004 (T-EMBLrel. 27, Last annotation update)  
 DE BAALC protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa;  
 OC Mammalia; Eutheria; Primata; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Ronaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC035038; AAH35038.1; -  
 DR EMBL; AF363578; AAL50378.1; -  
 DR InterPro; IPR009728; BAALC\_N.  
 DR Pfam; PF06989; BAALC\_N; 1.  
 SQ SEQUENCE 54 AA; 5663 MW; FEF0B4EABED9B829 CRC64;  
 Query Match 36.2%; Score 278; DB 2; Length 54;  
 Best Local Similarity 94.4%; Pred. No. 8.5e-19;  
 Matches 51; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIPRYESWTRETESTWLTYSDDLPSAAATDSGPEAGLHAG 54  
 Db 1 MCGGSRADAIPRYESWTRETESTWLTYSDDLPSAAATDSGPEAGLHAG 54  
 RESULT 13  
 Q8WXS2  
 ID Q8WXS2 PRELIMINARY; PRT; 80 AA.  
 AC Q8WXS2;  
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE BAALC 1-2.  
 DE Name=BAALC;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21574584; PubMed=11707601;  
 RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
 RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,  
 RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 RT lineage, is implicated in hematopoiesis and acute leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:13901-13906(2001).  
 DR EMBL; AF363578; AAL50382.1; -  
 DR InterPro; IPR009728; BAALC\_N; 1.  
 DR Pfam; PF06989; BAALC\_N; 1.  
 SQ SEQUENCE 80 AA; 8573 MW; 70BE9B91C3245E2D CRC64;  
 Query Match 35.5%; Score 272; DB 2; Length 80;  
 Best Local Similarity 94.3%; Pred. No. 5e-18;  
 Matches 50; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MCGGSRADAIPRYESWTRETESTWLTYSDDLPSAAATDSGPEAGLHA 53  
 Db 1 MCGGSRADAIPRYESWTRETESTWLTYSDDLPSAAATDSGPEAGLHA 53  
 RESULT 14  
 Q8WNE8  
 ID Q8WNE8 PRELIMINARY; PRT; 54 AA.  
 AC Q8WNE8;  
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE BAALC isoform 1-8.  
 GN Name=BAALC;  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21574584; PubMed=11707601;  
 RA Tanner S.M., Austin J.L., Leone G., Rush L.J., Plass C., Heinonen K.,  
 RA Mrozek K., Sill H., Knuutila S., Kolitz J.E., Archer K.J.,  
 RA Caligiuri M.A., Bloomfield C.D., de la Chapelle A.;  
 RT "BAALC, the human member of a novel mammalian neuroectoderm gene  
 RT lineage, is implicated in hematopoiesis and acute leukemia.";

Search completed: November 17, 2004, 15:27:52  
Job time : 105.667 secs

